GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: September 10, 2004, 10:37:15 ; Search time 763 Seconds
(without alignments)
11592.069 Million cell updates/sec

Title: US-10-076-069-1
Perfect score: 2082
Sequence: 1 ccccgcgtcggtcttccacc.....atggaaaaaaaaaaaaaaaa 2082

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2082	100.0	2082	6	AAL41497	Aal41497 Mouse hae
	2	428.4	20.6	2038	3	AAF18134	Aaf18134 Lung canc
	3	428.4	20.6	2038	4	AAH33113	Aah33113 Human col
	4	426.8	20.5	1314	6	ABL49854	Abl49854 Human CHD
	5	426.8	20.5	2082	6	AAL41498	Aal41498 Human hae
	6	425.2	20.4	2447	6	AAD34182	Aad34182 Human cDN
	7	425.2	20.4	2629	4	AAI60667	Aai60667 Human pol
	8	425.2	20.4	2629	4	AAI58881	Aai58881 Human pol
	9	425.2	20.4	2629	8	ADB48863	Adb48863 Novel hum
	10	355	17.1	862	4	AAH06538	Aah06538 Human cDN
	11	190	9.1	396	4	AAF94976	Aaf94976 Human ova
	12	190	9.1	396	6	ABL48926	Abl48926 Ovarian c
	13	190	9.1	396	6	ABT03243	Abt03243 Human ova
c	14	185.4	8.9	522	6	ABK63378	Abk63378 Rat seque
	15	98.2	4.7	477	8	ACH24999	Ach24999 Human adu
	16	94.8	4.6	1221	9	ADE07527	Ade07527 Novel cod
	17	62.4	3.0	368	6	ABL80236	Abl80236 Human ova
	18	61.2	2.9	18988	4	AAS46342	Aas46342 Tumour su
	19	61.2	2.9	18988	6	ABL32701	Abl32701 Human imm
	20	61.2	2.9	18988	6	ABL34509	Abl34509 Human met
	21	61.2	2.9	18988	6	ABL70204	Abl70204 Chemicall
	22	59	2.8	17721	6	ABL33728	Abl33728 Human imm
c	23	57.4	2.8	2000	7	ADA71938	Ada71938 Rice gene
	24	56.8	2.7	2000	7	ADA71938	Ada71938 Rice gene
	25	56.2	2.7	9642	6	ABL32356	Abl32356 Human imm
c	26	55.8	2.7	883	4	AAL15210	Aal15210 Human bre
	27	55.2	2.7	5689	4	AAS45384	Aas45384 Chemicall
	28	55.2	2.7	5689	4	AAS46426	Aas46426 Tumour su
	29	55.2	2.7	5689	6	ABK28226	Abk28226 DNA trans
	30	55.2	2.7	83391	6	ABQ67093	Abq67093 Human ang
	31	54.8	2.6	15548	6	ABL34155	Abl34155 Human imm
	32	54.6	2.6	5488	6	ABL33456	Abl33456 Human imm
	33	54.4	2.6	33053	6	ABQ67006	Abq67006 Human ang
	34	54.4	2.6	37973	6	ABL34196	Abl34196 Human imm
	35	54	2.6	8136	6	ABK39957	Abk39957 Human che
	36	54	2.6	8136	6	ABL32555	Abl32555 Human imm
	37	53.8	2.6	7589	6	ABK28390	Abk28390 DNA trans
	38	53.6	2.6	113515	6	ABL34174	Abl34174 Human imm
	39	53.4	2.6	6128	6	ABL32056	Abl32056 Human imm
	40	52.6	2.5	6904	6	ABL32212	Abl32212 Human imm
	41	52.2	2.5	5273	6	ABL32875	Abl32875 Human imm
	42	52	2.5	15161	6	ABL70457	Abl70457 Chemicall
	43	52	2.5	15161	6	AAS61422	Aas61422 Human gen
	44	51.8	2.5	6154	6	ABK31201	Abk31201 Signal tr
	45	51.8	2.5	6154	6	ABL70168	Abl70168 Chemicall

ALIGNMENTS

RESULT 1
AAL41497
ID AAL41497 standard; DNA; 2082 BP.
XX
AC AAL41497;
XX
DT 19-DEC-2002 (first entry)
XX
DE Mouse haematopoietic progenitor protein (Hepp) gene.
XX
KW Neuroprotective; nootropic; cytostatic; neurodegenerative disease; blood;
KW amyotrophic sclerosis; haematological disorder; neoplasm; leukaemia;
KW acute myelomonocytic leukaemia; lymphoblastic lymphoma; multiple myeloma;
KW chronic lymphocytic leukaemia; acute lymphoblastic leukaemia;
KW B-prolymphocytic leukaemia; plasma cell leukaemia; large B-cell lymphoma;
KW adult T-cell lymphoma; nodal marginal zone B-cell lymphoma; stem cell;
KW Burkitt's lymphoma; follicular lymphoma; hairy cell leukaemia;
KW mantle cell lymphoma; splenic marginal zone B-cell lymphoma;
KW T-prolymphocytic leukaemia; haematopoietic cytokine; growth factor;
KW progenitor cell; gene therapy; Hepp; haematopoietic progenitor protein;
KW mouse; gene; ds.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT CDS 191. .904
FT /*tag= a
FT /product= "Mouse Hepp protein"
FT polyA_signal 2045. .2049
FT /*tag= b
XX
PN WO200266610-A2.
XX
PD 29-AUG-2002.
XX
PF 15-FEB-2002; 2002WO-US004503.
XX
PR 16-FEB-2001; 2001US-0268923P.
XX
PA (UYMI-) UNIV MIAMI.
XX
PI Jurecic R, Nachtman RG;
XX
DR WPI; 2002-674928/72.
DR P-PSDB; AAO22897.
XX

PT New hematopoietic progenitor protein (Hepp) genes and proteins, useful
PT for detecting, treating and preventing neurodegenerative diseases, e.g.
PT amyotrophic sclerosis, and hematological disorders, e.g. neoplasms of the
PT blood.
XX
PS Claim 5; Fig 1A; 54pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising at least 85%
CC identity to either of 2 2082 base pair sequences, given in the
CC specification. The nucleic acids and polypeptides of the invention are
CC useful for detecting, treating and preventing neurodegenerative diseases
CC such as amyotrophic sclerosis, and haematological disorders, particularly
CC neoplasms of the blood such as acute myelomonocytic leukaemia,
CC lymphoblastic lymphoma, chronic lymphocytic leukaemia, acute
CC lymphoblastic leukaemia, multiple myeloma, B-prolymphocytic leukaemia,
CC plasma cell leukaemia, adult T-cell lymphoma/leukaemia, diffuse large B-
CC cell lymphoma, nodal marginal zone B-cell lymphoma, Burkitt's lymphoma,
CC follicular lymphoma, hairy cell leukaemia, mantle cell lymphoma, splenic
CC marginal zone B-cell lymphoma, and T-prolymphocytic leukaemia. They are
CC also useful as reagents for differential identification of tissues and
CC cell types present in the biological sample. The mammal is useful in
CC screening drugs for treating the disorders cited above, and for testing
CC of novel haematopoietic cytokines/growth factors for mobilisation and
CC differentiation of stem and progenitor cells. The nucleic acids of the
CC invention can be used in gene therapy. This polynucleotide sequence
CC represents the mouse haematopoietic progenitor protein (Hepp) gene of the
CC invention
XX
SQ Sequence 2082 BP; 484 A; 520 C; 522 G; 556 T; 0 U; 0 Other;

Query Match 100.0%; Score 2082; DB 6; Length 2082;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2082; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CCCCGCGTCGGTCTTCCACCTCACCTTTCGAGCTGGCCGCCGCTTGCTGTGCGCAGTTTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCCCGCGTCGGTCTTCCACCTCACCTTTCGAGCTGGCCGCCGCTTGCTGTGCGCAGTTTC 60

Qy         61 GGGGGACTGGACCTTCCCTGGCTTTTAGCAGCGCCGAGCGCCATGGCGACCCTTTGCTGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGGGGACTGGACCTTCCCTGGCTTTTAGCAGCGCCGAGCGCCATGGCGACCCTTTGCTGG 120

Qy        121 GCAGGTGACCGATTCCGGGTGCCCGAAGGAGCTGGCGTGGGTCTGCCTTGCAGCCGCCCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCAGGTGACCGATTCCGGGTGCCCGAAGGAGCTGGCGTGGGTCTGCCTTGCAGCCGCCCG 180

Qy        181 CCTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGAAGGAGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGAAGGAGT 240

Qy        241 AGAGGGTTTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACAGTCACTCCTGGACATGTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGAGGGTTTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACAGTCACTCCTGGACATGTC 300

Qy        301 CCTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAATCTCTGCCGCTCGGTCCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CCTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAATCTCTGCCGCTCGGTCCT 360

Qy        361 CATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCAGGATGGTGTGTGGCATGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCAGGATGGTGTGTGGCATGG 420

Qy        421 GATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCTGGTGTCCACAGAGATCCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCTGGTGTCCACAGAGATCCT 480

Qy        481 GTGTCGTACAGTGAGGGGAGCTGAGGAAGAGCACCCTGCTCCTGAACTGGAAGATGCTCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GTGTCGTACAGTGAGGGGAGCTGAGGAAGAGCACCCTGCTCCTGAACTGGAAGATGCTCC 540

Qy        541 CTTGCAAAACTCGGTTTCCGAGCTCCCCATCGTTGGCTCAGCACCAGGGCAAAGGAACCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CTTGCAAAACTCGGTTTCCGAGCTCCCCATCGTTGGCTCAGCACCAGGGCAAAGGAACCC 600

Qy        601 TCAGAGCAGCCTCTGGGAGATGGACAGCCCACAAGAAAACAGGGGAAGCTTTCAGAAGTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TCAGAGCAGCCTCTGGGAGATGGACAGCCCACAAGAAAACAGGGGAAGCTTTCAGAAGTC 660

Qy        661 ACTGGACCAGATATTTGAGACCCTGGAGAACAAAAACTCCAGTTCAGTGGAGGAACTCTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACTGGACCAGATATTTGAGACCCTGGAGAACAAAAACTCCAGTTCAGTGGAGGAACTCTT 720

Qy        721 CTCAGATGTGGACAGCTCCTACTATGACCTGGACACAGTGCTAACAGGAATGATGAGTGG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTCAGATGTGGACAGCTCCTACTATGACCTGGACACAGTGCTAACAGGAATGATGAGTGG 780

Qy        781 GACCAAGTCCAGTCTCTGCAATGGCCTTGAGGGCTTTGCTGCAGCCACCCCTCCTCCCAG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GACCAAGTCCAGTCTCTGCAATGGCCTTGAGGGCTTTGCTGCAGCCACCCCTCCTCCCAG 840

Qy        841 TTCCACTTGCAAGTCTGACCTGGCTGAGCTGGACCATGTGGTAGAGATTCTGGTGGAGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TTCCACTTGCAAGTCTGACCTGGCTGAGCTGGACCATGTGGTAGAGATTCTGGTGGAGAC 900

Qy        901 CTGAGAGGCCACCCCAGTGGGCTAAGGGTGAGGCCACCAGTCCCCATGGAGCTCACGTGT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTGAGAGGCCACCCCAGTGGGCTAAGGGTGAGGCCACCAGTCCCCATGGAGCTCACGTGT 960

Qy        961 GTTGTGACCCAGAGACAGATAAGCACTTGTCCTAAGAGGGGCTCTGGCTCTTGAGCTCAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTTGTGACCCAGAGACAGATAAGCACTTGTCCTAAGAGGGGCTCTGGCTCTTGAGCTCAT 1020

Qy       1021 TATCCTTTTGTGTGACATTGGACTCACTGTGGAGGATGGTGTGTCACAGCTATGTCTAGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TATCCTTTTGTGTGACATTGGACTCACTGTGGAGGATGGTGTGTCACAGCTATGTCTAGT 1080

Qy       1081 CTATTTTCAATTAGATAGGTGAACTTTCTAAAATTAAGTTTTATATGTTTTTGGGCAATA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CTATTTTCAATTAGATAGGTGAACTTTCTAAAATTAAGTTTTATATGTTTTTGGGCAATA 1140

Qy       1141 TTTTGTCTTAAGATATATTTTTTTAAACTTTTTATACTTTAGATTTTTTTCAGCTATTTTC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TTTTGTCTTAAGATATATTTTTTTAAACTTTTTATACTTTAGATTTTTTTCAGCTATTTTC 1200

Qy       1201 TTAAAAGTATATTTTTTCTACAAACATCCTCTGCTGCTACATTAGAAACATTTATAACCT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TTAAAAGTATATTTTTTCTACAAACATCCTCTGCTGCTACATTAGAAACATTTATAACCT 1260

Qy       1261 AAATACGATTGGTGTGTCATTTTAAAGGTTTAAATAGAAAACTTCTTTTTGTTACTGAGTC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAATACGATTGGTGTGTCATTTTAAAGGTTTAAATAGAAAACTTCTTTTTGTTACTGAGTC 1320

Qy       1321 TCTACACTCCCAAGGCAACTGTAAATGTAGCCGGCCGGGTGTTTACATGAGAGGCTCCAG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TCTACACTCCCAAGGCAACTGTAAATGTAGCCGGCCGGGTGTTTACATGAGAGGCTCCAG 1380

Qy       1381 TATGGTCTACATTCTAGTAGAGCTTGAAAAGAACCATGCACAGCTCCACTGCCCCCTCAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TATGGTCTACATTCTAGTAGAGCTTGAAAAGAACCATGCACAGCTCCACTGCCCCCTCAC 1440

Qy       1441 TGGGTCTGCTCTGGCGGATCGGAGCTCTCTTCCTAGCCCCGTGTGCAGGATGGCTTTATT 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TGGGTCTGCTCTGGCGGATCGGAGCTCTCTTCCTAGCCCCGTGTGCAGGATGGCTTTATT 1500

Qy       1501 TATGCCTATTTATATGTAAATGCCACTGAAAGCTAAGGTCTTACTCCTGGAAATCCCAAC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TATGCCTATTTATATGTAAATGCCACTGAAAGCTAAGGTCTTACTCCTGGAAATCCCAAC 1560

Qy       1561 ACCAGTTCTTCAGGGACTGCTGTGAGGCAGTGCCTTATGCAGGTCTTGTCCTTGGCCATC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 ACCAGTTCTTCAGGGACTGCTGTGAGGCAGTGCCTTATGCAGGTCTTGTCCTTGGCCATC 1620

Qy       1621 ACTGTCTGGTTCCCAGCCCAGCACATGTGACATGAGGACATGACATGCCCGAACCACCCA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 ACTGTCTGGTTCCCAGCCCAGCACATGTGACATGAGGACATGACATGCCCGAACCACCCA 1680
```

```
Qy       1681 GCACCACATGCTCCATGTCAAGTGTGTACGTGGAGACCACTGGCTCCCAGGCCTGTGCTC 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 GCACCACATGCTCCATGTCAAGTGTGTACGTGGAGACCACTGGCTCCCAGGCCTGTGCTC 1740

Qy       1741 AGAGAGGGTGTGCAGTCCTACGTGTGCTGGGGGGGACGACGGTGACCTGTGCTTGCTTGC 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 AGAGAGGGTGTGCAGTCCTACGTGTGCTGGGGGGGACGACGGTGACCTGTGCTTGCTTGC 1800

Qy       1801 TTTTAAAATGGTGCTTGGACGTTTTAAGGTTAAAAACAATCCGACTCCATATGATTTAGG 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 TTTTAAAATGGTGCTTGGACGTTTTAAGGTTAAAAACAATCCGACTCCATATGATTTAGG 1860

Qy       1861 GCTCCTCCACCCTGGGGTGGCCCCTATGCTGTCTGCTTGGATCTCAAAGTCTTGGTACTC 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GCTCCTCCACCCTGGGGTGGCCCCTATGCTGTCTGCTTGGATCTCAAAGTCTTGGTACTC 1920

Qy       1921 GGCACTGTCAGACTCCACCCCATGTATCCTTTTTGTTTCTCTTGTGCTTTTTTTGGACTT 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 GGCACTGTCAGACTCCACCCCATGTATCCTTTTTGTTTCTCTTGTGCTTTTTTTGGACTT 1980

Qy       1981 CCCAACCTGAGCCTAAGGTTTTATTTTATATGTGCTTCAATATCAACAATGTAAACCTCA 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 CCCAACCTGAGCCTAAGGTTTTATTTTATATGTGCTTCAATATCAACAATGTAAACCTCA 2040

Qy       2041 CTTTATTAAAAGTATCCAGCAAATGGAAAAAAAAAAAAAAAA 2082
             ||||||||||||||||||||||||||||||||||||||||||||
Db       2041 CTTTATTAAAAGTATCCAGCAAATGGAAAAAAAAAAAAAAAA 2082
```

RESULT 2
AAF18134

```
ID   AAF18134 standard; DNA; 2038 BP.
XX
AC   AAF18134;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Lung cancer associated polynucleotide sequence SEQ ID 153.
XX
KW   Human; lung cancer associated protein; neuroprotective; cytostatic;
KW   cardioactive; immunomodulatory; muscular active; vulnerary;
KW   gastrointestinal; nephrotropic; antiinfective; gynecological;
KW   antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW   proliferative disorder; wound healing; infectious disease; ds.
XX
OS   Homo sapiens.
XX
PN   WO200055180-A2.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US005918.
XX
PR   12-MAR-1999;   99US-0124270P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (ROSE/) ROSEN C A.
XX
PI   Ruben SM;
XX
DR   WPI; 2000-587514/55.
DR   P-PSDB; AAB58258.
XX
PT   Lung cancer associated gene sequences, referred to as lung cancer
PT   antigens, useful for treatment, prevention, and diagnosis of disorders
PT   such as lung cancer.
XX
PS   Claim 1; Page 618; 1425pp; English.
XX
CC   Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC   associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC   associated proteins and polynucleotide sequences, their agonists, and
CC   antagonists may have neuroprotective; cytostatic; cardioactive;
CC   immunomodulatory; muscular active general; vulnerary; gastrointestinal
CC   general; nephrotropic; antiinfective; gynecological; or antibacterial
CC   activity. The invention also includes antibodies specific for the protein
CC   or polynucleotide sequences. The lung cancer associated polynucleotide
CC   sequences may be used for detection of lung cancer, chromosome
CC   identification, as chromosome markers, and for numerous other diagnostic
CC   or research purposes. The proteins may be used to treat disorders such as
CC   neural, immune, muscular, reproductive, gastrointestinal, pulmonary,
CC   cardiovascular, renal, and proliferative disorders. The proteins may also
CC   be used in the treatment of wounds and infectious diseases.
CC   Polynucleotide sequences AAF18425 - AAF18433 and peptide AAB58549 are
CC   used in the course of the invention for the identification and
CC   characterisation of the polynucleotide and protein sequences
XX
SQ   Sequence 2038 BP; 439 A; 529 C; 552 G; 508 T; 0 U; 10 Other;
```

Query Match 20.6%; Score 428.4; DB 3; Length 2038;
Best Local Similarity 74.3%; Pred. No. 9.6e-93;
Matches 559; Conservative 0; Mismatches 181; Indels 12; Gaps 1;

```
Qy        175 CGCCCGCCTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGA 234
              |||  | | |||||  ||||||||  |||| ||||||||||||| || || ||| || ||
Db         82 CGCGGGTCAGGACACAATGTTTGCACGAGGACTGAAGAGGAAATGTGTTGGCCACGAGGA 141

Qy        235 AGGAGTAGAGGGT-----------TTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACA 282
              ||  || |||||              ||    || ||  | || || ||||||||||| ||
Db        142 AGACGTGGAGGGAGCCCTGGCCGGCTTGAAGACAGTGTCCTCATACAGCCTGCAGCGGCA 201

Qy        283 GTCACTCCTGGACATGTCCCTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAA 342
              ||| |||||||||||||| || || ||| | ||||| || |||||||| ||||||||||
Db        202 GTCGCTCCTGGACATGTCTCTGGTGAAGTTGCAGCTTTGCCACATGCTTGTGGAGCCCAA 261

Qy        343 TCTCTGCCGCTCGGTCCTCATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCA 402
              ||| |||||||| |||||||| |||||||| |||||||||||||| ||||| ||||  ||
Db        262 TCTGTGCCGCTCAGTCCTCATTGCCAACACGGTCCGGCAGATCCAAGAGGAGATGACGCA 321

Qy        403 GGATGGTGTGTGGCATGGGATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCT 462
              ||||||   |||||       ||||||||||  || ||| ||||| ||  | || ||| |
Db        322 GGATGGGACGTGGCGCACAGTGGCACCCCAGGCTGCAGAGCGGGCGCCGYTCGACCGCTT 381

Qy        463 GGTGTCCACAGAGATCCTGTGTCGTACAGTGAGGGGAGCTGAGGAAGAGCACCCTGCTCC 522
              ||| ||||| |||||||||| ||| ||| | ||||    ||||  |  || ||||||||
Db        382 GGTCTCCACGGAGATCCTGTGCCGTGCAGCGTGGGGGCAAGAGGGGGCACATCCTGCTCC 441

Qy        523 TGAACTGGAAGATGCTCCCTTGCAAAACTCGGTTTCCGAGCTCCCCATCGTTGGCTCAGC 582
              ||   |||  || |  | |   ||      | |||||  || ||   |   ||   ||||||
Db        442 TGGCTTGGGGGACGGCCACACACAGGGTCCAGTTTCTGACCTTTGCCCAGTCACCTCAGC 501

Qy        583 ACCAGGGCAAAGGAACCCTCAGAGCAGCCTCTGGGAGATGGACAGCCCACAAGAAAACAG 642
              ||  |  | |||| |||  |||||||||  ||||||||||  |||| | |||||||||
Db        502 ACAGGCACCAAGGCACCTGCAGAGCAGCGCCTGGGAGATGGATGGCCCTCGAGAAAACAG 561

Qy        643 GGGAAGCTTTCAGAAGTCACTGGACCAGATATTTGAGACCCTGGAGAACAAAAACTCCAG 702
              ||||||||||| |||||||| || ||||||||||| || |||||||  |||||| ||||
Db        562 AGGAAGCTTTCACAAGTCACTTGATCAGATATTTGAAACGCTGGAGACTAAAAACCCCAG 621

Qy        703 TTCAGTGGAGGAACTCTTCTCAGATGTGGACAGCTCCTACTATGACCTGGACACAGTGCT 762
              |   |||| || || |||||||| ||||||||| ||||||| |||||||||||||| ||
Db        622 CTGCATGGAAGAGCTGTTCTCAGACGTGGACAGCCCCTACTACGACCTGGACACAGTACT 681

Qy        763 AACAGGAATGATGAGTGGGACCAAGTCCAGTCTCTGCAATGGCCTTGAGGGCTTTGCTGC 822
              ||||| |||||| | ||  ||| | |  | | |||| | || || ||||||| ||| |
Db        682 GACAGGCATGATGGGGGGTGCCAGGCCGGGCCCCTGCGAAGGGCTCGAGGGCTTGGCTCC 741

Qy        823 AGCCACCCCTCCTCCCAGTTCCACTTGCAAGTCTGACCTGGCTGAGCTGGACCATGTGGT 882
              ||||||||    || || ||||  |||||||| |||||||  ||||||||||| |||||
Db        742 GGCCACCCCRGGCCCTAGCTCCAGCTGCAAGTCCGACCTGGGCGAGCTGGACCACGTGGT 801

Qy        883 AGAGATTCTGGTGGAGACCTGAGAGGCCACCC 914
              ||||| ||||||||||||||||  |   |||
Db        802 GGAGATCCTGGTGGAGACCTGAGCAGGAGCCC 833
```

RESULT 3
AAH33113
ID AAH33113 standard; cDNA; 2038 BP.
XX
AC AAH33113;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen encoding cDNA SEQ ID NO:169.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; ss.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US026524.
XX
PR 29-SEP-1999; 99US-0157137P.
PR 03-NOV-1999; 99US-0163280P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI; 2001-235357/24.
DR P-PSDB; AAG73682.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers.
XX
PS Claim 1; Page 2326; 9803pp; English.
XX
CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where the
CC proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene therapy
CC and vaccine production. N and P may be used in the prevention, diagnosis
CC and treatment of diseases associated with inappropriate P expression. For
CC example, N and P may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of P by expressing inactive proteins or to
CC supplement the patients own production of P. Additionally, N may be used
CC to produce the colon cancer-associated Ps, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the proteins. N and P
CC can be used in the prevention, diagnosis and treatment of colorectal
CC carcinomas and cancers. AAH37196 to AAH37204 and AAB77789 represent
CC sequences used in the exemplification of the present invention. N.B.
CC Pages 666 to 682 and page 7053 of the sequence listing were missing at
CC time of publication, meaning no sequences are present for SEQ ID NO:1027
CC to 1052, 7921 and 7922
XX
SQ Sequence 2038 BP; 439 A; 529 C; 552 G; 508 T; 0 U; 10 Other;

Query Match 20.6%; Score 428.4; DB 4; Length 2038;
Best Local Similarity 74.3%; Pred. No. 9.6e-93;
Matches 559; Conservative 0; Mismatches 181; Indels 12; Gaps 1;

```
Qy      175 CGCCCGCCTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGA 234
            |||  | | |||||  ||||||||  |||| ||||||||||||| || || ||| || ||
Db       82 CGCGGGTCAGGACACAATGTTTGCACGAGGACTGAAGAGGAAATGTGTTGGCCACGAGGA 141

Qy      235 AGGAGTAGAGGGT----------TTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACA 282
            ||  || |||||           ||    || ||  | || || ||||||||||| ||
Db      142 AGACGTGGAGGGAGCCCTGGCCGGCTTGAAGACAGTGTCCTCATACAGCCTGCAGCGGCA 201

Qy      283 GTCACTCCTGGACATGTCCCTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAA 342
            ||| |||||||||||||| || || ||| | ||||| || |||||||| ||||||||||
Db      202 GTCGCTCCTGGACATGTCTCTGGTGAAGTTGCAGCTTTGCCACATGCTTGTGGAGCCCAA 261

Qy      343 TCTCTGCCGCTCGGTCCTCATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCA 402
            ||| |||||||| |||||||| |||||||| |||||||||||||| ||||| ||||  ||
Db      262 TCTGTGCCGCTCAGTCCTCATTGCCAACACGGTCCGGCAGATCCAAGAGGAGATGACGCA 321

Qy      403 GGATGGTGTGTGGCATGGGATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCT 462
            ||||||   |||||      |||||||||||  || ||| ||||| ||  | || ||| |
Db      322 GGATGGGACGTGGCGCACAGTGGCACCCCAGGCTGCAGAGCGGGCGCCGYTCGACCGCTT 381

Qy      463 GGTGTCCACAGAGATCCTGTGTCGTACAGTGAGGGGAGCTGAGGAAGAGCACCCTGCTCC 522
            ||| ||||| |||||||||||| ||| ||| | ||||    ||||  |  || ||||||||
Db      382 GGTCTCCACGGAGATCCTGTGCCGTGCAGCGTGGGGGCAAGAGGGGGCACATCCTGCTCC 441

Qy      523 TGAACTGGAAGATGCTCCCTTGCAAAACTCGGTTTCCGAGCTCCCCATCGTTGGCTCAGC 582
            ||   |||  || |  | |   ||      | ||||| || ||    |   ||   ||||||
Db      442 TGGCTTGGGGGACGGCCACACACAGGGTCCAGTTTCTGACCTTTGCCCAGTCACCTCAGC 501

Qy      583 ACCAGGGCAAAGGAACCCTCAGAGCAGCCTCTGGGAGATGGACAGCCCACAAGAAAACAG 642
            ||  |  | |||| |||  |||||||||  ||||||||||||  |||| | |||||||||
Db      502 ACAGGCACCAAGGCACCTGCAGAGCAGCGCCTGGGAGATGGATGGCCCTCGAGAAAACAG 561

Qy      643 GGGAAGCTTTCAGAAGTCACTGGACCAGATATTTGAGACCCTGGAGAACAAAAACTCCAG 702
            |||||||||||  |||||||| || ||||||||||| || |||||||  |||||| ||||
Db      562 AGGAAGCTTTCACAAGTCACTTGATCAGATATTTGAAACGCTGGAGACTAAAAACCCCAG 621

Qy      703 TTCAGTGGAGGAACTCTTCTCAGATGTGGACAGCTCCTACTATGACCTGGACACAGTGCT 762
            |   |||| || || |||||||| |||||||| ||||||| |||||||||||||| ||
Db      622 CTGCATGGAAGAGCTGTTCTCAGACGTGGACAGCCCCTACTACGACCTGGACACAGTACT 681

Qy      763 AACAGGAATGATGAGTGGGACCAAGTCCAGTCTCTGCAATGGCCTTGAGGGCTTTGCTGC 822
            ||||| |||||| | ||  ||| | |  | |||| | || || |||||||| ||| |
Db      682 GACAGGCATGATGGGGGGTGCCAGGCCGGGCCCCTGCGAAGGGCTCGAGGGCTTGGCTCC 741

Qy      823 AGCCACCCCTCCTCCCAGTTCCACTTGCAAGTCTGACCTGGCTGAGCTGGACCATGTGGT 882
            ||||||||    || || ||||  |||||||| |||||||  |||||||||| |||||
Db      742 GGCCACCCCRGGCCCTAGCTCCAGCTGCAAGTCCGACCTGGGCGAGCTGGACCACGTGGT 801

Qy      883 AGAGATTCTGGTGGAGACCTGAGAGGCCACCC 914
            ||||| ||||||||||||||||  |   |||
Db      802 GGAGATCCTGGTGGAGACCTGAGCAGGAGCCC 833
```

RESULT 4
ABL49854
ID ABL49854 standard; cDNA; 1314 BP.
XX
AC ABL49854;
XX
DT 05-JUN-2002 (first entry)
XX
DE Human CHD protein 26.51 encoding cDNA SEQ ID NO:1.
XX
KW Human; CHD protein 26.51; malignant tumour; haemopathy; HIV infection;
KW development disturbance; immunological disease; inflammation; gene; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 98. .823
FT /*tag= a
FT /product= "Human CHD protein 26.51"
XX
PN CN1324856-A.
XX
PD 05-DEC-2001.
XX
PF 24-MAY-2000; 2000CN-00115839.
XX
PR 24-MAY-2000; 2000CN-00115839.
XX
PA (BODE-) BODE GENE DEV CO LTD SHANGHAI.
XX
PI Mao Y, Xie Y;
XX

```
DR   WPI; 2002-217513/28.
DR   P-PSDB; ABB06375.
XX
PT   New polypeptide human CHD protein 26.51 and polynucleotides for encoding
PT   same.
XX
PS   Claim 6; Page 25-26 (Disclosure); 33pp; Chinese.
XX
CC   The present sequence encodes human CHD protein 26.51 (I). The present
CC   invention also describes a method for producing (I) using DNA
CC   recombination techniques. (I) and the polynucleotide encoding it can be
CC   used in the treatment of several diseases, such as malignant tumour,
CC   haemopathy, development disturbances, HIV infection, immunological
CC   diseases and various inflammations
XX
SQ   Sequence 1314 BP; 264 A; 376 C; 386 G; 288 T; 0 U; 0 Other;

  Query Match             20.5%;  Score 426.8;  DB 6;  Length 1314;
  Best Local Similarity   74.2%;  Pred. No. 1.9e-92;
  Matches  558;  Conservative    0;  Mismatches  182;  Indels   12;  Gaps    1;

Qy        175 CGCCCGCCTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGA 234
              |||  | | |||||  ||||||||  |||| ||||||||||||| || || ||| || ||
Db         82 CGCGGGTCAGGACACAATGTTTGCACGAGGACTGAAGAGGAAATGTGTTGGCCACGAGGA 141

Qy        235 AGGAGTAGAGGGT------------TTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACA 282
              ||  || |||||            ||    || ||  | || || |||||||||| ||
Db        142 AGACGTGGAGGGAGCCCTGGCCGGCTTGAAGACAGTGTCCTCATACAGCCTGCAGCGGCA 201

Qy        283 GTCACTCCTGGACATGTCCCTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAA 342
              ||| |||||||||||||| || || ||| | ||||| || |||||||| ||||||||||
Db        202 GTCGCTCCTGGACATGTCTCTGGTGAAGTTGCAGCTTTGCCACATGCTTGTGGAGCCCAA 261

Qy        343 TCTCTGCCGCTCGGTCCTCATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCA 402
              || |||||||| |||||||| |||||||| |||||||||||||| ||||| ||||  ||
Db        262 CCTGTGCCGCTCAGTCCTCATTGCCAACACGGTCCGGCAGATCCAAGAGGAGATGACGCA 321

Qy        403 GGATGGTGTGTGGCATGGGATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCT 462
              ||||||   |||||       ||||||||||  || ||| ||||| ||  | || ||| |
Db        322 GGATGGGACGTGGCGCACAGTGGCACCCCAGGCTGCAGAGCGGGCGCCGCTCGACCGCTT 381

Qy        463 GGTGTCCACAGAGATCCTGTGTCGTACAGTGAGGGGAGCTGAGGAAGAGCACCCTGCTCC 522
              ||| ||||| |||||||||||| ||| ||| | ||||    ||||  |  || ||||||||
Db        382 GGTCTCCACGGAGATCCTGTGCCGTGCAGCGTGGGGGCAAGAGGGGGCACATCCTGCTCC 441

Qy        523 TGAACTGGAAGATGCTCCCTTGCAAAACTCGGTTTCCGAGCTCCCCATCGTTGGCTCAGC 582
              ||   |||  || |  | |   ||      | ||||| || ||   |   ||   ||||||
Db        442 TGGCTTGGGGGACGGCCACACACAGGGTCCAGTTTCTGACCTTTGCCCAGTCACCTCAGC 501

Qy        583 ACCAGGGCAAAGGAACCCTCAGAGCAGCCTCTGGGAGATGGACAGCCCACAAGAAAACAG 642
              ||  |  | |||| |||  |||||||||  ||||||||||||  |||| | |||||||||
Db        502 ACAGGCACCAAGGCACCTGCAGAGCAGCGCCTGGGAGATGGATGGCCCTCGAGAAAACAG 561

Qy        643 GGGAAGCTTTCAGAAGTCACTGGACCAGATATTTGAGACCCTGGAGAACAAAAACTCCAG 702
              |||||||||||| |||||||| || |||||||||||  || |||||||  |||||| ||||
Db        562 AGGAAGCTTTCACAAGTCACTTGATCAGATATTTGAAACGCTGGAGACTAAAAACCCCAG 621

Qy        703 TTCAGTGGAGGAACTCTTCTCAGATGTGGACAGCTCCTACTATGACCTGGACACAGTGCT 762
              |   |||| || || |||||||| |||||||||  |||||||  ||||||||||||||| ||
Db        622 CTGCATGGAAGAGCTGTTCTCAGACGTGGACAGCCCCTACTACGACCTGGACACAGTACT 681

Qy        763 AACAGGAATGATGAGTGGGACCAAGTCCAGTCTCTGCAATGGCCTTGAGGGCTTTGCTGC 822
              ||||| |||||| | ||  ||| | |  | | |||| | || || ||||||||  ||| |
Db        682 GACAGGCATGATGGGGGGTGCCAGGCCGGGCCCCTGCGAAGGGCTCGAGGGCTTGGCTCC 741

Qy        823 AGCCACCCCTCCTCCCAGTTCCACTTGCAAGTCTGACCTGGCTGAGCTGGACCATGTGGT 882
              ||||||||     || || ||||  |||||||| |||||||  |||||||||| |||||
Db        742 GGCCACCCCAGGCCCTAGCTCCAGCTGCAAGTCCGACCTGGGCGAGCTGGACCACGTGGT 801

Qy        883 AGAGATTCTGGTGGAGACCTGAGAGGCCACCC 914
              |||||  ||||||||||||||||  |   |||
Db        802 GGAGATCCTGGTGGAGACCTGAGCAGGAGCCC 833
```

```
RESULT 5
AAL41498
ID   AAL41498 standard; DNA; 2082 BP.
XX
AC   AAL41498;
XX
DT   19-DEC-2002  (first entry)
XX
DE   Human haematopoietic progenitor protein (Hepp) gene.
XX
KW   Neuroprotective; nootropic; cytostatic; neurodegenerative disease; blood;
KW   amyotrophic sclerosis; haematological disorder; neoplasm; leukaemia;
KW   acute myelomonocytic leukaemia; lymphoblastic lymphoma; multiple myeloma;
KW   chronic lymphocytic leukaemia; acute lymphoblastic leukaemia;
KW   B-prolymphocytic leukaemia; plasma cell leukaemia; large B-cell lymphoma;
KW   adult T-cell lymphoma; nodal marginal zone B-cell lymphoma; stem cell;
KW   Burkitt's lymphoma; follicular lymphoma; hairy cell leukaemia;
KW   mantle cell lymphoma; splenic marginal zone B-cell lymphoma;
KW   T-prolymphocytic leukaemia; haematopoietic cytokine; growth factor;
KW   progenitor cell; gene therapy; Hepp; haematopoietic progenitor protein;
KW   human; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             75. .800
FT                   /*tag=  a
FT                   /product= "Human Hepp protein"
FT   polyA_signal    2043. .2047
FT                   /*tag=  b
XX
PN   WO200266610-A2.
XX
PD   29-AUG-2002.
XX
PF   15-FEB-2002; 2002WO-US004503.
XX
PR   16-FEB-2001; 2001US-0268923P.
XX
PA   (UYMI-) UNIV MIAMI.
XX
PI   Jurecic R,  Nachtman RG;
XX
DR   WPI; 2002-674928/72.
DR   P-PSDB; AAO22898.
XX
PT   New hematopoietic progenitor protein (Hepp) genes and proteins, useful
PT   for detecting, treating and preventing neurodegenerative diseases, e.g.
PT   amyotrophic sclerosis, and hematological disorders, e.g. neoplasms of the
PT   blood.
XX
PS   Claim 5; Fig 1B; 54pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising at least 85%
CC   identity to either of 2 2082 base pair sequences, given in the
CC   specification. The nucleic acids and polypeptides of the invention are
CC   useful for detecting, treating and preventing neurodegenerative diseases
CC   such as amyotrophic sclerosis, and haematological disorders, particularly
CC   neoplasms of the blood such as acute myelomonocytic leukaemia,
CC   lymphoblastic lymphoma, chronic lymphocytic leukaemia, acute
CC   lymphoblastic leukaemia, multiple myeloma, B-prolymphocytic leukaemia,
CC   plasma cell leukaemia, adult T-cell lymphoma/leukaemia, diffuse large B-
CC   cell lymphoma, nodal marginal zone B-cell lymphoma, Burkitt's lymphoma,
CC   follicular lymphoma, hairy cell leukaemia, mantle cell lymphoma, splenic
CC   marginal zone B-cell lymphoma, and T-prolymphocytic leukaemia. They are
CC   also useful as reagents for differential identification of tissues and
CC   cell types present in the biological sample. The mammal is useful in
CC   screening drugs for treating the disorders cited above, and for testing
CC   of novel haematopoietic cytokines/growth factors for mobilisation and
CC   differentiation of stem and progenitor cells. The nucleic acids of the
CC   invention can be used in gene therapy. This polynucleotide sequence
CC   represents the human haematopoietic progenitor protein (Hepp) gene of the
```

```
CC   invention
XX
SQ   Sequence 2082 BP; 471 A; 539 C; 545 G; 527 T; 0 U; 0 Other;

  Query Match             20.5%;  Score 426.8;  DB 6;  Length 2082;
  Best Local Similarity   74.2%;  Pred. No. 2.4e-92;
  Matches  558;  Conservative    0;  Mismatches  182;  Indels   12;  Gaps    1;

Qy        175 CGCCCGCCTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGA 234
              |||   | | |||||  ||||||||  ||||  |||||||||||||| || || ||| || ||
Db         59 CGCGGGTCAGGACACAATGTTTGCACGAGGACTGAAGAGGAAATGTGTTGGCCACGAGGA 118

Qy        235 AGGAGTAGAGGGT-----------TTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACA 282
              ||  || |||||             ||    || ||  | || || ||||||||||| ||
Db        119 AGACGTGGAGGGAGCCCTGGCCGGCTTGAAGACAGTGTCCTCATACAGCCTGCAGCGGCA 178

Qy        283 GTCACTCCTGGACATGTCCCTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAA 342
              ||| ||||||||||||||| || || ||| | ||||| || |||||||| ||||||||||
Db        179 GTCGCTCCTGGACATGTCTCTGGTGAAGTTGCAGCTTTGCCACATGCTTGTGGAGCCCAA 238

Qy        343 TCTCTGCCGCTCGGTCCTCATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCA 402
              || |||||||| |||||||| |||||||| |||||||||||||| ||||| ||||  ||
Db        239 CCTGTGCCGCTCAGTCCTCATTGCCAACACGGTCCGGCAGATCCAAGAGGAGATGACGCA 298

Qy        403 GGATGGTGTGTGGCATGGGATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCT 462
              ||||||   |||||      |||||||||||  || ||| ||||| ||  | || ||| |
Db        299 GGATGGGACGTGGCGCACAGTGGCACCCCAGGCTGCAGAGCGGGCGCCGCTCGACCGCTT 358

Qy        463 GGTGTCCACAGAGATCCTGTGTCGTACAGTGAGGGGAGCTGAGGAAGAGCACCCTGCTCC 522
              ||| ||||| |||||||||||| ||| ||| | ||||     |||| |  || ||||||||
Db        359 GGTCTCCACGGAGATCCTGTGCCGTGCAGCGTGGGGGCAAGAGGGGGCACATCCTGCTCC 418

Qy        523 TGAACTGGAAGATGCTCCCTTGCAAAACTCGGTTTCCGAGCTCCCCATCGTTGGCTCAGC 582
              ||   |||  || |  | |    ||      | ||||| || ||   |   ||   ||||||
Db        419 TGGCTTGGGGGACGGCCACACACAGGGTCCAGTTTCTGACCTTTGCCCAGTCACCTCAGC 478

Qy        583 ACCAGGGCAAAGGAACCCTCAGAGCAGCCTCTGGGAGATGGACAGCCCACAAGAAAACAG 642
              ||  |  | |||| |||  |||||||||  ||||||||||||  |||| | |||||||||
Db        479 ACAGGCACCAAGGCACCTGCAGAGCAGCGCCTGGGAGATGGATGGCCCTCGAGAAAACAG 538

Qy        643 GGGAAGCTTTCAGAAGTCACTGGACCAGATATTTGAGACCCTGGAGAACAAAAACTCCAG 702
              |||||||||||| |||||||| || |||||||||||  || |||||||  ||||||  ||||
Db        539 AGGAAGCTTTCACAAGTCACTTGATCAGATATTTGAAACGCTGGAGACTAAAAACCCCAG 598

Qy        703 TTCAGTGGAGGAACTCTTCTCAGATGTGGACAGCTCCTACTATGACCTGGACACAGTGCT 762
              |   |||| || || |||||||| |||||||||| |||||||  ||||||||||||||| ||
Db        599 CTGCATGGAAGAGCTGTTCTCAGACGTGGACAGCCCCTACTACGACCTGGACACAGTACT 658

Qy        763 AACAGGAATGATGAGTGGGACCAAGTCCAGTCTCTGCAATGGCCTTGAGGGCTTTGCTGC 822
              |||||  ||||||  | ||  |||  | |   | | ||||  | ||  || ||||||||  ||| |
Db        659 GACAGGCATGATGGGGGGTGCCAGGCCGGGCCCCTGCGAAGGGCTCGAGGGCTTGGCTCC 718

Qy        823 AGCCACCCCTCCTCCCAGTTCCACTTGCAAGTCTGACCTGGCTGAGCTGGACCATGTGGT 882
              |||||||||    || || ||||  |||||||| |||||||  ||||||||||| |||||
Db        719 GGCCACCCCAGGCCCTAGCTCCAGCTGCAAGTCCGACCTGGGCGAGCTGGACCACGTGGT 778

Qy        883 AGAGATTCTGGTGGAGACCTGAGAGGCCACCC 914
              ||||| ||||||||||||||||   |   |||
Db        779 GGAGATCCTGGTGGAGACCTGAGCAGGAGCCC 810


RESULT 6
AAD34182
ID   AAD34182 standard; cDNA; 2447 BP.
XX
AC   AAD34182;
XX
DT   16-JUL-2002  (first entry)
XX
DE   Human cDNA coexpressed along with cyclin B and UbcH10 cell cycle gene.
XX
KW   Human; cell cycle gene; topo II; PRC1; CDC23; cyclin B; UbcH10; hp55cdc;
KW   cancer; immune disorder; adenocarcinoma; leukaemia; lymphoma; melanoma;
KW   myeloma; sarcoma; atherosclerosis; Crohn's disease; glomerulonephritis;
KW   multiple sclerosis; asthma; myasthenia gravis; rheumatoid arthritis;
KW   scleroderma; osteoporosis; systemic lupus erythematosus; cytostatic;
KW   nephrotropic; neuroprotective; dermatological; immunosuppressive;
KW   osteopathic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200218575-A2.
XX
PD   07-MAR-2002.
XX
PF   27-AUG-2001; 2001WO-US026682.
XX
PR   30-AUG-2000; 2000US-0229253P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Walker MG,  Jung K;
XX
DR   WPI; 2002-329773/36.
XX
PT   Composition comprising cDNA molecules coexpressed with one or more known
PT   cell cycle genes, useful for diagnosis and treatment of cell cycle
PT   disorders e.g. glomerulonephritis, multiple sclerosis, rheumatoid
PT   arthritis.
XX
PS   Claim 6; Page 38-39; 43pp; English.
XX
CC   The present invention relates to compositions comprising cDNA molecules
CC   coexpressed with one or more known cell cycle genes such as topo II,
CC   PRC1, CDC23, cyclin B, UbcH10 and hp55cdc. The composition is useful for
CC   screening several molecules or compounds such as DNA or RNA molecules,
CC   peptide nucleic acids, enhancers, transcription factors, repressors,
CC   mimetics or proteins, identify an purify a ligand. Sequences of the
CC   invention are useful for treating cell cycle disorders such as cancer or
CC   immune disorder such as adenocarcinoma, leukaemia, lymphoma, melanoma,
CC   myeloma, sarcoma or cancers of the blood or bone, asthma,
CC   atherosclerosis, Crohn's disease, glomerulonephritis, multiple sclerosis,
CC   myasthenia gravis, osteoporosis, rheumatoid arthritis, scleroderma or
CC   systemic lupus erythematosus. They are useful as surrogate markers in
CC   diagnosis, prognosis and evaluation of therapies for cell cycle disorders
CC   and potentially serve as the therapeutic for the elimination or control
CC   of unregulated cell cycling. The proteins or peptides expressed from the
CC   cDNAs of the invention are either potential therapeutics or targets for
CC   identification or development of therapeutics. The present sequence is a
CC   cDNA coexpressed along with cyclin B and UbcH10 cell cycle gene. This
CC   cDNA is used as a specific diagnostic marker for breast cancer
XX
SQ   Sequence 2447 BP; 538 A; 623 C; 638 G; 648 T; 0 U; 0 Other;

  Query Match             20.4%;  Score 425.2;  DB 6;  Length 2447;
  Best Local Similarity   74.1%;  Pred. No. 6.1e-92;
  Matches  557;  Conservative    0;  Mismatches  183;  Indels   12;  Gaps    1;

Qy        175 CGCCCGCCTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGA 234
              |||   | | |||||  ||||||||  ||||  |||||||||||||| || || ||| || ||
Db         75 CGCGGGTCAGGACACAATGTTTGCACGAGGACTGAAGAGGAAATGTGTTGGCCACGAGGA 134

Qy        235 AGGAGTAGAGGGT-----------TTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACA 282
              ||  || |||||             ||    || ||  | || || ||||||||||| ||
Db        135 AGACGTGGAGGGAGCCCTGGCCGGCTTGAAGACAGTGTCCTCATACAGCCTGCAGCGGCA 194

Qy        283 GTCACTCCTGGACATGTCCCTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAA 342
              ||| ||||||||||||||| || || ||| | ||||| || |||||||| ||||||||||
Db        195 GTCGCTCCTGGACATGTCTCTGGTGAAGTTGCAGCTTTGCCACATGCTTGTGGAGCCCAA 254

Qy        343 TCTCTGCCGCTCGGTCCTCATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCA 402
              || |||||||| |||||||| |||||||| |||||||||||||| ||||| ||||  ||
Db        255 CCTGTGCCGCTCAGTCCTCATTGCCAACACGGTCCGGCAGATCCAAGAGGAGATGACGCA 314

Qy        403 GGATGGTGTGTGGCATGGGATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCT 462
              ||||||   |||||      |||||||||||  || ||| ||||| ||  | || ||| |
```

```
Db        315 GGATGGGACGTGGCGCACAGTGGCACCCCAGGCTGCAGAGCGGGCGCCGCTCGACCGCTT 374

Qy        463 GGTGTCCACAGAGATCCTGTGTCGTACAGTGAGGGGAGCTGAGGAAGAGCACCCTGCTCC 522

Db        375 GGTCTCCACGGAGATCCTGTGCCGTGCAGCGTGGGGGCAAGAGGGGGCACATCCTGCTCC 434

Qy        523 TGAACTGGAAGATGCTCCCTTGCAAAACTCGGTTTCCGAGCTCCCCATCGTTGGCTCAGC 582

Db        435 TGGCTTGGGGGACGGCCACACACAGGGTCCAGTTTCTGACCTTTGCCCAGTCACCTCAGC 494

Qy        583 ACCAGGGCAAAGGAACCCTCAGAGCAGCCTCTGGGAGATGGACAGCCCACAAGAAAACAG 642

Db        495 ACAGGCACCAAGGCACCTGCAGAGCAGCGCCTGGGAGATGGATGGCCCTCGAGAAAACAG 554

Qy        643 GGGAAGCTTTCAGAAGTCACTGGACCAGATATTTGAGACCCTGGAGAACAAAAACTCCAG 702

Db        555 AGGAAGCTTTCACAAGTCACTTGATCAGATATTTGAAACGCTGGAGACTAAAAACCCCAG 614

Qy        703 TTCAGTGGAGGAACTCTTCTCAGATGTGGACAGCTCCTACTATGACCTGGACACAGTGCT 762

Db        615 CTGCATGGAAGAGCTGTTCTCAGACGTGGACAGCCCCTACTACGACCTGGACACAGTACT 674

Qy        763 AACAGGAATGATGAGTGGGACCAAGTCCAGTCTCTGCAATGGCCTTGAGGGCTTTGCTGC 822

Db        675 GACAGGCATGATGGGGGGTGCCAGGCCGGGCCCCTGCGAAGGGCTCGAGGGCTTGGCTCC 734

Qy        823 AGCCACCCCTCCTCCCAGTTCCACTTGCAAGTCTGACCTGGCTGAGCTGGACCATGTGGT 882

Db        735 GGCCACCCCAGGCCCTAGCTCCAGCTGCAAGTCCGACCTGGGCGAGCTGGACCACGTGAT 794

Qy        883 AGAGATTCTGGTGGAGACCTGAGAGGCCACCC 914

Db        795 GGAGATCCTGGTGGAGACCTGAGCAGGAGCCC 826
```

RESULT 7
AAI60667
ID AAI60667 standard; cDNA; 2629 BP.
XX
AC AAI60667;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 4656.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US034263.
XX
PR 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR P-PSDB; AAM41511.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.
XX
PS Claim 1; SEQ ID NO 4656; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX
SQ Sequence 2629 BP; 581 A; 652 C; 691 G; 705 T; 0 U; 0 Other;

Query Match 20.4%; Score 425.2; DB 4; Length 2629;
Best Local Similarity 74.1%; Pred. No. 6.3e-92;
Matches 557; Conservative 0; Mismatches 183; Indels 12; Gaps 1;

```
Qy        175 CGCCCGCCTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGA 234

Db         57 CGCGGGTCAGGACACAATGTTTGCACGAGGACTGAAGAGGAAATGTGTTGGCCACGAGGA 116

Qy        235 AGGAGTAGAGGGT-----------TTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACA 282

Db        117 AGACGTGGAGGGAGCCCTGGCCGGCTTGAAGACAGTGTCCTCATACAGCCTGCAGCGGCA 176

Qy        283 GTCACTCCTGGACATGTCCCTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAA 342

Db        177 GTCGCTCCTGGACATGTCTCTGGTGAAGTTGCAGCTTTGCCACATGCTTGTGGAGCCCAA 236

Qy        343 TCTCTGCCGCTCGGTCCTCATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCA 402

Db        237 CCTGTGCCGCTCAGTCCTCATTGCCAACACGGTCCGGCAGATCCAAGAGGAGATGACGCA 296

Qy        403 GGATGGTGTGTGGCATGGGATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCT 462

Db        297 GGATGGGACGTGGCGCACAGTGGCACCCCAGGCTGCAGAGCGGGCGCCGCTCGACCGCTT 356

Qy        463 GGTGTCCACAGAGATCCTGTGTCGTACAGTGAGGGGAGCTGAGGAAGAGCACCCTGCTCC 522

Db        357 GGTCTCCACGGAGATCCTGTGCCGTGCAGCGTGGGGGCAAGAGGGGGCACATCCTGCTTC 416

Qy        523 TGAACTGGAAGATGCTCCCTTGCAAAACTCGGTTTCCGAGCTCCCCATCGTTGGCTCAGC 582

Db        417 TGGCTTGGGGGACGGCCACACACAGGGTCCAGTTTCTGACCTTTGCCCAGTCACCTCAGC 476

Qy        583 ACCAGGGCAAAGGAACCCTCAGAGCAGCCTCTGGGAGATGGACAGCCCACAAGAAAACAG 642

Db        477 ACAGGCACCAAGGCACCTGCAGAGCAGCGCCTGGGAGATGGATGGCCCTCGAGAAAACAG 536

Qy        643 GGGAAGCTTTCAGAAGTCACTGGACCAGATATTTGAGACCCTGGAGAACAAAAACTCCAG 702

Db        537 AGGAAGCTTTCACAAGTCACTTGATCAGATATTTGAAACGCTGGAGACTAAAAACCCCAG 596

Qy        703 TTCAGTGGAGGAACTCTTCTCAGATGTGGACAGCTCCTACTATGACCTGGACACAGTGCT 762

Db        597 CTGCATGGAAGAGCTGTTCTCAGACGTGGACAGCCCCTACTACGACCTGGACACAGTACT 656
```

```
Qy        763 AACAGGAATGATGAGTGGGACCAAGTCCAGTCTCTGCAATGGCCTTGAGGGCTTTGCTGC 822
              ||||| ||||||  | ||  ||| | |  | | |||| | || || |||||||| ||| |
Db        657 GACAGGCATGATGGGGGGTGCCAGGCCGGGCCCCTGCGAAGGGCTCGAGGGCTTGGCTCC 716

Qy        823 AGCCACCCCTCCTCCCAGTTCCACTTGCAAGTCTGACCTGGCTGAGCTGGACCATGTGGT 882
              ||||||||    || || ||||  |||||||| |||||||  |||||||||| |||||
Db        717 GGCCACCCCAGGCCCTAGCTCCAGCTGCAAGTCCGACCTGGGCGAGCTGGACCACGTGGT 776

Qy        883 AGAGATTCTGGTGGAGACCTGAGAGGCCACCC 914
              ||||| ||||||||||||||||  |   |||
Db        777 GGAGATCCTGGTGGAGACCTGAGCAGGAGCCC 808
```

RESULT 8
AAI58881
ID AAI58881 standard; cDNA; 2629 BP.
XX
AC AAI58881;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 1084.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US034263.
XX
PR 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR P-PSDB; AAM39725.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.
XX
PS Claim 1; SEQ ID NO 1084; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX
SQ Sequence 2629 BP; 581 A; 652 C; 691 G; 705 T; 0 U; 0 Other;

```
  Query Match           20.4%;  Score 425.2;  DB 4;  Length 2629;
  Best Local Similarity  74.1%;  Pred. No. 6.3e-92;
  Matches  557;  Conservative    0;  Mismatches  183;  Indels   12;  Gaps    1;

Qy        175 CGCCCGCCTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGA 234
              |||  | | |||||  |||||||  |||| ||||||||||||| || || ||| || ||
Db         57 CGCGGGTCAGGACACAATGTTTGCACGAGGACTGAAGAGGAAATGTGTTGGCCACGAGGA 116

Qy        235 AGGAGTAGAGGGT-----------TTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACA 282
              ||  || |||||           ||    || ||  | || || |||||||||||| ||
Db        117 AGACGTGGAGGGAGCCCTGGCCGGCTTGAAGACAGTGTCCTCATACAGCCTGCAGCGGCA 176

Qy        283 GTCACTCCTGGACATGTCCCTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAA 342
              ||| ||||||||||||||| || || ||| | ||||| || |||||||| ||||||||||
Db        177 GTCGCTCCTGGACATGTCTCTGGTGAAGTTGCAGCTTTGCCACATGCTTGTGGAGCCCAA 236

Qy        343 TCTCTGCCGCTCGGTCCTCATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCA 402
              || |||||||| |||||||| |||||||| |||||||||||||| ||||| ||||  ||
Db        237 CCTGTGCCGCTCAGTCCTCATTGCCAACACGGTCCGGCAGATCCAAGAGGAGATGACGCA 296

Qy        403 GGATGGTGTGTGGCATGGGATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCT 462
              ||||||  |||||      |||||||||||  || ||| |||||  ||  | || ||| |
Db        297 GGATGGGACGTGGCGCACAGTGGCACCCCAGGCTGCAGAGCGGGCGCCGCTCGACCGCTT 356

Qy        463 GGTGTCCACAGAGATCCTGTGTCGTACAGTGAGGGGAGCTGAGGAAGAGCACCCTGCTCC 522
              ||| ||||| |||||||||||| ||| ||| | ||||     ||||  |  || ||||||| |
Db        357 GGTCTCCACGGAGATCCTGTGCCGTGCAGCGTGGGGGCAAGAGGGGGCACATCCTGCTTC 416

Qy        523 TGAACTGGAAGATGCTCCCTTGCAAAACTCGGTTTCCGAGCTCCCCATCGTTGGCTCAGC 582
              ||   |||  || |  | |    ||      | ||||| || ||    |   ||   ||||||
Db        417 TGGCTTGGGGGACGGCCACACACAGGGTCCAGTTTCTGACCTTTGCCCAGTCACCTCAGC 476

Qy        583 ACCAGGGCAAAGGAACCCTCAGAGCAGCCTCTGGGAGATGGACAGCCCACAAGAAAACAG 642
              ||  |  | |||| |||  |||||||||  ||||||||||||  |||| | |||||||||
Db        477 ACAGGCACCAAGGCACCTGCAGAGCAGCGCCTGGGAGATGGATGGCCCTCGAGAAAACAG 536

Qy        643 GGGAAGCTTTCAGAAGTCACTGGACCAGATATTTGAGACCCTGGAGAACAAAAACTCCAG 702
              |||||||||| |||||||| || |||||||||| || |||||||  ||||||  ||||
Db        537 AGGAAGCTTTCACAAGTCACTTGATCAGATATTTGAAACGCTGGAGACTAAAAACCCCAG 596

Qy        703 TTCAGTGGAGGAACTCTTCTCAGATGTGGACAGCTCCTACTATGACCTGGACACAGTGCT 762
              |   |||| || || |||||||| |||||||| ||||||| ||||||||||||||| ||
Db        597 CTGCATGGAAGAGCTGTTCTCAGACGTGGACAGCCCCTACTACGACCTGGACACAGTACT 656

Qy        763 AACAGGAATGATGAGTGGGACCAAGTCCAGTCTCTGCAATGGCCTTGAGGGCTTTGCTGC 822
              ||||| ||||||  | ||  ||| | |  | | |||| | || || |||||||| ||| |
Db        657 GACAGGCATGATGGGGGGTGCCAGGCCGGGCCCCTGCGAAGGGCTCGAGGGCTTGGCTCC 716

Qy        823 AGCCACCCCTCCTCCCAGTTCCACTTGCAAGTCTGACCTGGCTGAGCTGGACCATGTGGT 882
              ||||||||    || || ||||  |||||||| |||||||  |||||||||| |||||
Db        717 GGCCACCCCAGGCCCTAGCTCCAGCTGCAAGTCCGACCTGGGCGAGCTGGACCACGTGGT 776

Qy        883 AGAGATTCTGGTGGAGACCTGAGAGGCCACCC 914
              ||||| ||||||||||||||||  |   |||
Db        777 GGAGATCCTGGTGGAGACCTGAGCAGGAGCCC 808
```

RESULT 9
ADB48863
ID ADB48863 standard; cDNA; 2629 BP.
XX
AC ADB48863;
XX
DT 04-DEC-2003 (first entry)
XX
DE Novel human cDNA SEQ ID NO 773.

```
XX
KW   ss; cancer; neurodegenerative disease; human.
XX
OS   Homo sapiens.
XX
PN   US2003104529-A1.
XX
PD   05-JUN-2003.
XX
PF   04-JAN-2002; 2002US-00037270.
XX
PR   21-JAN-2000; 2000US-00488725.
PR   25-APR-2000; 2000US-00552317.
PR   19-JUL-2000; 2000US-00620312.
XX
PA   (ZHOU/) ZHOU P.
PA   (TANG/) TANG Y T.
PA   (LIUC/) LIU C.
PA   (ASUN/) ASUNDI V.
PA   (DRMA/) DRMANAC R T.
XX
PI   Zhou P,  Tang YT,  Liu C,  Asundi V,  Drmanac RT;
XX
DR   WPI; 2003-678194/64.
XX
PT   New polynucleotide, useful for treating diseases e.g., cancer or
PT   neurodegenerative diseases.
XX
PS   Claim 1; SEQ ID NO 773; 99pp; English.
XX
CC   The invention relates to a polynucleotide comprising a sequence given in
CC   the specification, or its mature protein-coding portion, or its
CC   complement. The polynucleotide is useful for treating diseases e.g.,
CC   cancer or neurodegenerative diseases and many others listed in the
CC   specification. The present sequence represents a novel human cDNA. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format directly from USPTO
CC   at seqdata.uspto.gov/sequence.html?DocID=20030104529.
XX
SQ   Sequence 2629 BP; 581 A; 652 C; 691 G; 705 T; 0 U; 0 Other;

  Query Match             20.4%;  Score 425.2;  DB 8;  Length 2629;
  Best Local Similarity   74.1%;  Pred. No. 6.3e-92;
  Matches  557;  Conservative    0;  Mismatches  183;  Indels   12;  Gaps    1;

Qy        175 CGCCCGCCTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGA 234
              |||  | | |||||  ||||||||  |||| ||||||||||||| || || ||| || ||
Db         57 CGCGGGTCAGGACACAATGTTTGCACGAGGACTGAAGAGGAAATGTGTTGGCCACGAGGA 116

Qy        235 AGGAGTAGAGGGT-----------TTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACA 282
              ||  || |||||           ||    || ||  | || || ||||||||||| ||
Db        117 AGACGTGGAGGGAGCCCTGGCCGGCTTGAAGACAGTGTCCTCATACAGCCTGCAGCGGCA 176

Qy        283 GTCACTCCTGGACATGTCCCTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAA 342
              ||| ||||||||||||||| || || ||| | ||||| || |||||||| ||||||||||
Db        177 GTCGCTCCTGGACATGTCTCTGGTGAAGTTGCAGCTTTGCCACATGCTTGTGGAGCCCAA 236

Qy        343 TCTCTGCCGCTCGGTCCTCATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCA 402
               || |||||||| |||||||| |||||||| |||||||||||||| ||||| ||||   ||
Db        237 CCTGTGCCGCTCAGTCCTCATTGCCAACACGGTCCGGCAGATCCAAGAGGAGATGACGCA 296

Qy        403 GGATGGTGTGTGGCATGGGATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCT 462
              ||||||   |||||      ||||||||||  || ||| ||||| ||  | || ||| |
Db        297 GGATGGGACGTGGCGCACAGTGGCACCCCAGGCTGCAGAGCGGGCGCCGCTCGACCGCTT 356

Qy        463 GGTGTCCACAGAGATCCTGTGTCGTACAGTGAGGGGAGCTGAGGAAGAGCACCCTGCTCC 522
              ||| ||||| |||||||||||| ||| ||| | ||||    ||||  |  || |||||| |
Db        357 GGTCTCCACGGAGATCCTGTGCCGTGCAGCGTGGGGGCAAGAGGGGGCACATCCTGCTTC 416

Qy        523 TGAACTGGAAGATGCTCCCTTGCAAAACTCGGTTTCCGAGCTCCCCATCGTTGGCTCAGC 582
              ||   |||  || |  | |    ||      | ||||| || ||   |   ||   ||||||
Db        417 TGGCTTGGGGGACGGCCACACACAGGGTCCAGTTTCTGACCTTTGCCCAGTCACCTCAGC 476

Qy        583 ACCAGGGCAAAGGAACCCTCAGAGCAGCCTCTGGGAGATGGACAGCCCACAAGAAAACAG 642
              ||  |  | |||| |||  |||||||||  ||||||||||||  |||| | |||||||||
Db        477 ACAGGCACCAAGGCACCTGCAGAGCAGCGCCTGGGAGATGGATGGCCCTCGAGAAAACAG 536

Qy        643 GGGAAGCTTTCAGAAGTCACTGGACCAGATATTTGAGACCCTGGAGAACAAAAACTCCAG 702
              ||||||||||| |||||||| || |||||||||| || |||||||  |||||| ||||
Db        537 AGGAAGCTTTCACAAGTCACTTGATCAGATATTTGAAACGCTGGAGACTAAAAACCCCAG 596

Qy        703 TTCAGTGGAGGAACTCTTCTCAGATGTGGACAGCTCCTACTATGACCTGGACACAGTGCT 762
              |   |||| || || |||||||| ||||||||| ||||||| |||||||||||||| ||
Db        597 CTGCATGGAAGAGCTGTTCTCAGACGTGGACAGCCCCTACTACGACCTGGACACAGTACT 656

Qy        763 AACAGGAATGATGAGTGGGACCAAGTCCAGTCTCTGCAATGGCCTTGAGGGCTTTGCTGC 822
              ||||| |||||| | ||  ||| | |  | | |||| | || || |||||||| ||| |
Db        657 GACAGGCATGATGGGGGGTGCCAGGCCGGGCCCCTGCGAAGGGCTCGAGGGCTTGGCTCC 716

Qy        823 AGCCACCCCTCCTCCCAGTTCCACTTGCAAGTCTGACCTGGCTGAGCTGGACCATGTGGT 882
              ||||||||    || || |||| |||||||| |||||||  |||||||||| |||||
Db        717 GGCCACCCCAGGCCCTAGCTCCAGCTGCAAGTCCGACCTGGGCGAGCTGGACCACGTGGT 776

Qy        883 AGAGATTCTGGTGGAGACCTGAGAGGCCACCC 914
              ||||| ||||||||||||||||  |   |||
Db        777 GGAGATCCTGGTGGAGACCTGAGCAGGAGCCC 808


RESULT 10
AAH06538
ID   AAH06538 standard; cDNA; 862 BP.
XX
AC   AAH06538;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA clone (5'-primer) SEQ ID NO:3373.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-00116126.
XX
PR   29-JUL-1999;   99JP-00248036.
PR   27-AUG-1999;   99JP-00300253.
PR   11-JAN-2000; 2000JP-00118776.
PR   02-MAY-2000; 2000JP-00183767.
PR   09-JUN-2000; 2000JP-00241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT   length cDNAs defined in the specification, and for the detection and/or
PT   diagnosis of the abnormality of the proteins encoded by the full-length
PT   cDNAs.
XX
PS   Claim 1; SEQ ID NO 3373; 2537pp + Sequence Listing; English.
XX
CC   The present invention describes primer sets for synthesising 5602 full-
CC   length cDNAs defined in the specification. Where a primer set comprises:
CC   (a) an oligo-dT primer and an oligonucleotide complementary to the
CC   complementary strand of a polynucleotide which comprises one of the 5602
CC   nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
```

```
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in the
CC   specification. The primer sets can be used in antisense therapy and in
CC   gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC   represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC   oligonucleotides, all of which are used in the exemplification of the
CC   present invention
XX
SQ   Sequence 862 BP; 183 A; 241 C; 286 G; 147 T; 0 U; 5 Other;

  Query Match             17.1%;  Score 355;  DB 4;  Length 862;
  Best Local Similarity   73.7%;  Pred. No. 3.2e-75;
  Matches  484;  Conservative    0;  Mismatches  160;  Indels   13;  Gaps    2;

Qy        175 CGCCCGCCTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGA 234
              ||| | | |||||  ||||||||  |||| ||||||||||||| || || ||| || ||
Db         65 CGCGGGTCAGGACACAATGTTTGCACGAGGACTGAAGAGGAAATGTGTTGGCCACGAGGA 124

Qy        235 AGGAGTAGAGGGT-----------TTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACA 282
              ||  || |||||           ||    || ||  | || || |||||||||| ||
Db        125 AGACGTGGAGGGAGCCCTGGCCGGCTTGAAGACAGTGTCCTCATACAGCCTGCAGCGGCA 184

Qy        283 GTCACTCCTGGACATGTCCCTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAA 342
              ||| |||||||||||||| || || ||| | ||||| || |||||||| ||||||||||
Db        185 GTCGCTCCTGGACATGTCTCTGGTGAAGTTGCAGCTTTGCCACATGCTTGTGGAGCCCAA 244

Qy        343 TCTCTGCCGCTCGGTCCTCATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCA 402
              || |||||||| |||||||| |||||||| |||||||||||||| ||||| ||||  ||
Db        245 CCTGTGCCGCTCAGTCCTCATTGCCAACACGGTCCGGCAGATCCAAGAGGAGATGACGCA 304

Qy        403 GGATGGTGTGTGGCATGGGATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCT 462
              ||||||   |||||      |||||||||||  || ||| ||||| ||  | || ||| |
Db        305 GGATGGGACGTGGCGCACAGTGGCACCCCAGGCTGCAGAGCGGGCGCCGCTCGACCGCTT 364

Qy        463 GGTGTCCACAGAGATCCTGTGTCGTACAGTGAGGGGAGCTGAGGAAGAGCACCCTGCTCC 522
              ||| ||||| |||||||||||| ||| ||| | ||||    ||||  |  || |||||||
Db        365 GGTCTCCACGGAGATCCTGTGCCGTGCAGCGTGGGGGCAAGAGGGGGCACATCCTGCTCC 424

Qy        523 TGAACTGGAAGATGCTCCCTTGCAAAACTCGGTTTCCGAGCTCCCCATCGTTGGCTCAGC 582
              ||   |||  || |  | |   ||      | ||||| || ||    |   ||  ||||||
Db        425 TGGCTTGGGGGACGGCCACACACAGGGTCCAGTTTCTGACCTTTGCCCAGTCACCTCAGC 484

Qy        583 ACCAGGGCAAAGGAACCCTCAGAGCAGCCTCTGGGAGATGGACAGCCCACAAGAAAACAG 642
              ||  |  | |||| |||  |||||||||  ||||||||||||  |||| | |||||||||
Db        485 ACAGGCACCAAGGCACCTGCAGAGCAGCGCCTGGGAGATGGATGGCCCTCGAGAAAACAG 544

Qy        643 GGGAAGCTTTCAGAAGTCACTGGACCAGATATTTGAGACCCTGGAGAACAAAAACTCCAG 702
              |||||||||||  |||||||| || ||||||||||| || |||||||  |||||| ||||
Db        545 AGGAAGCTTTCACAAGTCACTTGATCAGATATTTGAAACGCTGGAGACTAAAAACCCCAG 604

Qy        703 TTCAGTGGAGGAACTCTTCTCAGATGTGGACAGCTCCTACTATGACCTGGACACAGTGC- 761
              |   |||| || || |||||||| |||||||| ||||||| ||||||||||||||| |
Db        605 CTGCATGGAAGAGCTGTTCTCAGACGTGGACAGCCCCTACTACGACCTGGACACAGTACT 664

Qy        762 TAACAGGAATGATGAGTGGGACCAAGTCCAGTCTCTGCAATGGCCTTGAGGGCTTTG 818
              | ||||| |||||| | ||  ||| | |  | | |||| | || || ||||||||||
Db        665 TGACAGGCATGATGGGGGGTGCCAGGCCGGGCCCCTGCGAAGGGCTCGAGGGCTTTG 721


RESULT 11
AAF94976
ID   AAF94976 standard; cDNA; 396 BP.
XX
AC   AAF94976;
XX
DT   23-MAY-2001  (first entry)
XX
DE   Human ovarian cancer associated coding sequence SEQ ID NO: 167.
XX
KW   Human, ovarian cancer; vaccine; gene therapy; carcinoma; ss.
XX
OS   Homo sapiens.
XX
PN   WO200118046-A2.
XX
PD   15-MAR-2001.
XX
PF   08-SEP-2000; 2000WO-US024827.
XX
PR   10-SEP-1999;   99US-00394374.
PR   01-MAY-2000; 2000US-00561778.
PR   15-AUG-2000; 2000US-00640173.
PR   07-SEP-2000; 2000US-00656668.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J,  Stolk JA;
XX
DR   WPI; 2001-211395/21.
XX
PT   Isolated polypeptides associated with ovarian carcinomas, and the nucleic
PT   acids that encode them, useful for the prevention diagnosis and treatment
PT   of ovarian cancers.
XX
PS   Claim 5; Page 170; 189pp; English.
XX
CC   The present invention provides a number of coding sequences and proteins,
CC   the over-expression of which is associated with ovarian carcinoma/cancer.
CC   These can be used in the diagnosis, treatment and prevention of ovarian
CC   cancer, optionally by gene therapy or in the form of a vaccine. The
CC   present sequence is an example of one of these sequences
XX
SQ   Sequence 396 BP; 75 A; 108 C; 147 G; 66 T; 0 U; 0 Other;

  Query Match              9.1%;  Score 190;  DB 4;  Length 396;
  Best Local Similarity   74.7%;  Pred. No. 1.3e-35;
  Matches  257;  Conservative    0;  Mismatches   75;  Indels   12;  Gaps    1;

Qy        175 CGCCCGCCTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGA 234
              ||| | | |||||  ||||||||  |||| ||||||||||||| || || ||| || ||
Db         52 CGCGGGTCAGGACACAATGTTTGCACGAGGACTGAAGAGGAAATGTGTTGGCCACGAGGA 111

Qy        235 AGGAGTAGAGGGT-----------TTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACA 282
              ||  || |||||           ||    || ||  | || || |||||||||| ||
Db        112 AGACGTGGAGGGAGCCCTGGCCGGCTTGAAGACAGTGTCCTCATACAGCCTGCAGCGGCA 171

Qy        283 GTCACTCCTGGACATGTCCCTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAA 342
              ||| |||||||||||||| || || ||| | ||||| || |||||||| ||||||||||
Db        172 GTCGCTCCTGGACATGTCTCTGGTGAAGTTGCAGCTTTGCCACATGCTTGTGGAGCCCAA 231

Qy        343 TCTCTGCCGCTCGGTCCTCATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCA 402
              ||| |||||||| |||||||| |||||||| |||||||||||||| ||||| ||||  ||
Db        232 TCTGTGCCGCTCAGTCCTCATTGCCAACACGGTCCGGCAGATCCAAGAGGAGATGACGCA 291

Qy        403 GGATGGTGTGTGGCATGGGATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCT 462
              ||||||   |||||      |||||||||||  || ||| ||||| ||  | || ||| |
Db        292 GGATGGGACGTGGCGCACAGTGGCACCCCAGGCTGCAGAGCGGGCGCCGCTCGACCGCTT 351

Qy        463 GGTGTCCACAGAGATCCTGTGTCGTACAGTGAGGGGAGCTGAGG 506
              ||| ||||| |||||||||||  ||| ||| | ||||    ||||
Db        352 GGTCTCCACGGAGATCCTGTGCCGTGCAGCGTGGGGGCAAGAGG 395


RESULT 12
ABL48926
ID   ABL48926 standard; cDNA; 396 BP.
XX
AC   ABL48926;
XX
```

```
DT   18-JUN-2002  (first entry)
XX
DE   Ovarian carcinoma sequence isolate 25796.
XX
KW   Ovarian cancer; cancer therapy; vaccine; gene therapy; tumour; cancer;
KW   ss.
XX
OS   Homo sapiens.
XX
PN   US2002004491-A1.
XX
PD   10-JAN-2002.
XX
PF   03-APR-2001; 2001US-00825294.
XX
PR   10-SEP-1999;   99US-00394374.
PR   01-MAY-2000; 2000US-00561778.
PR   15-AUG-2000; 2000US-00640173.
PR   07-SEP-2000; 2000US-00656668.
PR   14-NOV-2000; 2000US-00713550.
XX
PA   (XUJJ/) XU J.
PA   (STOL/) STOLK J A.
PA   (ALGA/) ALGATE P A.
PA   (FLIN/) FLING S P.
XX
PI   Xu J,  Stolk JA,  Algate PA,  Fling SP;
XX
DR   WPI; 2002-171027/22.
XX
PT   Ovarian tumor polypeptide and polynucleotide useful in diagnosis,
PT   prevention and/or treatment of cancer, especially ovarian cancer.
XX
PS   Claim 1a; Page 94; 131pp; English.
XX
CC   The invention relates to ovarian tumour polynucleotides and polypeptides
CC   that may be utilised in cancer therapy, for example in a vaccine or gene
CC   therapy. Polypeptides and polynucleotides of the invention are useful for
CC   detecting a cancer in a patient, for stimulating and/or expanding T-cells
CC   specific for a tumour protein, and for inhibiting the development of a
CC   cancer in a patient. They are also useful for stimulating an immune
CC   response in a patient, and for treating a cancer in a patient and for
CC   determining the presence of a cancer in a patient. The isolated
CC   polynucleotides of the invention are useful for their ability to
CC   selectively form duplex molecules with complementary stretches of the
CC   entire desired gene or gene fragments, and for designing and preparing
CC   ribozyme molecules for inhibiting expression of tumour polypeptides in
CC   tumour cells. Polypeptides and polynucleotides of the invention are also
CC   useful in recombinant DNA molecules to direct expression of a polypeptide
CC   in appropriate host cells. The sequences given in records ABL48760-
CC   ABL48956 represent polynucleotides encoding ovarian carcinoma proteins
XX
SQ   Sequence 396 BP; 75 A; 108 C; 147 G; 66 T; 0 U; 0 Other;

  Query Match             9.1%;  Score 190;  DB 6;  Length 396;
  Best Local Similarity   74.7%;  Pred. No. 1.3e-35;
  Matches  257;  Conservative    0;  Mismatches   75;  Indels   12;  Gaps    1;

Qy        175 CGCCCGCCTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGA 234
              |||  | | |||||  ||||||||  ||||| ||||||||||||| || || ||| || ||
Db         52 CGCGGGTCAGGACACAATGTTTGCACGAGGACTGAAGAGGAAATGTGTTGGCCACGAGGA 111

Qy        235 AGGAGTAGAGGGT------------TTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACA 282
              ||  || |||||              ||    || ||  | || || |||||||||||| ||
Db        112 AGACGTGGAGGGAGCCCTGGCCGGCTTGAAGACAGTGTCCTCATACAGCCTGCAGCGGCA 171

Qy        283 GTCACTCCTGGACATGTCCCTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAA 342
              ||| |||||||||||||| || || ||| | ||||| || |||||||| ||||||||||
Db        172 GTCGCTCCTGGACATGTCTCTGGTGAAGTTGCAGCTTTGCCACATGCTTGTGGAGCCCAA 231

Qy        343 TCTCTGCCGCTCGGTCCTCATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCA 402
              ||| |||||||| |||||||| |||||||| |||||||||||||| ||||| ||||  ||
Db        232 TCTGTGCCGCTCAGTCCTCATTGCCAACACGGTCCGGCAGATCCAAGAGGAGATGACGCA 291

Qy        403 GGATGGTGTGTGGCATGGGATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCT 462
              ||||||   |||||       |||||||||||  || ||| ||||| ||  | || ||| |
Db        292 GGATGGGACGTGGCGCACAGTGGCACCCCAGGCTGCAGAGCGGGCGCCGCTCGACCGCTT 351

Qy        463 GGTGTCCACAGAGATCCTGTGTCGTACAGTGAGGGGAGCTGAGG 506
              ||| ||||| |||||||||||| ||| ||| | ||||    ||||
Db        352 GGTCTCCACGGAGATCCTGTGCCGTGCAGCGTGGGGGCAAGAGG 395


RESULT 13
ABT03243
ID   ABT03243 standard; cDNA; 396 BP.
XX
AC   ABT03243;
XX
DT   05-SEP-2002  (first entry)
XX
DE   Human ovarian carcinoma associated coding sequence SEQ ID NO: 167.
XX
KW   Human; ovarian cancer; ovarian carcinoma; gene therapy; immunotherapy;
KW   cytostatic; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200239885-A2.
XX
PD   23-MAY-2002.
XX
PF   13-NOV-2001; 2001WO-US045395.
XX
PR   14-NOV-2000; 2000US-00713550.
PR   03-APR-2001; 2001US-00825294.
PR   02-OCT-2001; 2001US-00970966.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J,  Stolk JA,  Algate PA,  Fling SP,  Molesh DA;
XX
DR   WPI; 2002-500186/53.
XX
PT   Novel ovarian cancer polypeptide and polynucleotide, useful for detecting
PT   the presence of ovarian cancer in a patient, and in pharmaceutical
PT   compositions, e.g. vaccines, for treating ovarian cancer.
XX
PS   Example 1; Page 169-170; 197pp; English.
XX
CC   The present invention provides human ovarian cancer associated proteins
CC   and coding sequences. The sequences can be used in the diagnosis and
CC   treatment of ovarian cancers. The present sequence is a coding sequence
CC   of the invention
XX
SQ   Sequence 396 BP; 75 A; 108 C; 147 G; 66 T; 0 U; 0 Other;

  Query Match             9.1%;  Score 190;  DB 6;  Length 396;
  Best Local Similarity   74.7%;  Pred. No. 1.3e-35;
  Matches  257;  Conservative    0;  Mismatches   75;  Indels   12;  Gaps    1;

Qy        175 CGCCCGCCTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGA 234
              |||  | | |||||  ||||||||  ||||| ||||||||||||| || || ||| || ||
Db         52 CGCGGGTCAGGACACAATGTTTGCACGAGGACTGAAGAGGAAATGTGTTGGCCACGAGGA 111

Qy        235 AGGAGTAGAGGGT------------TTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACA 282
              ||  || |||||              ||    || ||  | || || |||||||||||| ||
Db        112 AGACGTGGAGGGAGCCCTGGCCGGCTTGAAGACAGTGTCCTCATACAGCCTGCAGCGGCA 171

Qy        283 GTCACTCCTGGACATGTCCCTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAA 342
              ||| |||||||||||||| || || ||| | ||||| || |||||||| ||||||||||
Db        172 GTCGCTCCTGGACATGTCTCTGGTGAAGTTGCAGCTTTGCCACATGCTTGTGGAGCCCAA 231

Qy        343 TCTCTGCCGCTCGGTCCTCATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCA 402
              ||| |||||||| |||||||| |||||||| |||||||||||||| ||||| ||||  ||
Db        232 TCTGTGCCGCTCAGTCCTCATTGCCAACACGGTCCGGCAGATCCAAGAGGAGATGACGCA 291
```

```
Qy        403 GGATGGTGTGTGGCATGGGATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCT 462
              ||||||  |||||     ||||||||||  || ||| ||||| ||  | || ||| |
Db        292 GGATGGGACGTGGCGCACAGTGGCACCCCAGGCTGCAGAGCGGGCGCCGCTCGACCGCTT 351

Qy        463 GGTGTCCACAGAGATCCTGTGTCGTACAGTGAGGGGAGCTGAGG 506
              ||| ||||| |||||||||||| ||| ||| | ||||    ||||
Db        352 GGTCTCCACGGAGATCCTGTGCCGTGCAGCGTGGGGGCAAGAGG 395
```

RESULT 14
ABK63378/c
ID ABK63378 standard; cDNA; 522 BP.
XX
AC ABK63378;
XX
DT 18-JUN-2002 (first entry)
XX
DE Rat sequence differentially expressed in response to a hepatotoxin #1285.
XX
KW Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;
KW differential expression; centrilobular necrosis; steatosis.
XX
OS Rattus norvegicus.
XX
PN WO200210453-A2.
XX
PD 07-FEB-2002.
XX
PF 30-JUL-2001; 2001WO-US023872.
XX
PR 31-JUL-2000; 2000US-0222040P.
PR 02-NOV-2000; 2000US-0244880P.
PR 11-MAY-2001; 2001US-0290029P.
PR 15-MAY-2001; 2001US-0290645P.
PR 22-MAY-2001; 2001US-0292336P.
PR 06-JUN-2001; 2001US-0295798P.
PR 13-JUN-2001; 2001US-0297457P.
PR 19-JUN-2001; 2001US-0298884P.
PR 09-JUL-2001; 2001US-0303459P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Mendrick D, Porter MW, Johnson KR, Castle AL, Elashoff MR;
XX
DR WPI; 2002-241625/29.
XX
PT Predicting toxic effects of compounds or the progression of these toxic
PT effects by determining the changes in gene expression in tissues or cells
PT exposed to the toxin and comparing these to gene expression in unexposed
PT tissues or cells.
XX
PS Claim 1; SEQ ID NO 1285; 239pp; English.
XX
CC The invention relates to methods for predicting toxic effects of
CC compounds or the progression of these toxic effects by determining the
CC global changes in gene expression in tissues or cells exposed to the
CC toxin and comparing these to gene expression in unexposed tissues or
CC cells. Also included are methods of predicting at least one toxic effect
CC of a compound or progression of a toxic effect, preferably the
CC hepatotoxicity of a compound, comprising detecting the level of
CC expression in a tissue or cell sample exposed to the compound of two or
CC more genes listed in the specification, where differential expression of
CC the genes is indicative of at least one toxic effect or progression. The
CC method can also be used to identify an agent which modulates the toxic
CC response and predict cellular pathways that a compound modulates in a
CC cell. The methods utilise a set of at least two probes (on a solid
CC support in kit form), where each of the probes comprises a sequence that
CC specifically hybridises to a gene listed in the specification, a computer
CC system comprising a database containing information identifying the
CC expression level in a tissue or cell sample exposed to a hepatotoxin of a
CC set of genes comprising at least two genes listed in the specification,
CC and a user interface to view the information used to present information
CC identifying the expression level in a tissue or cell of at least one gene
CC listed in the specification. The method is useful for elucidating global
CC changes in gene expression and for identifying toxicity markers in
CC tissues or cell exposed to a known toxin. The genes may be used as
CC toxicity markers in drug screening and toxicity assays. The genes and
CC gene expression information may be used as diagnostic markers for the
CC prediction or identification of the physiological state of tissue or cell
CC sample that has been exposed to a compound or agent. Hepatotoxicity is
CC characterised by centrilobular necrosis and steatosis. The present
CC sequence is an expressed sequence tag (EST) or cDNA derived from a gene
CC which is differentially expressed in response to a hepatotoxic agent
XX
SQ Sequence 522 BP; 149 A; 120 C; 136 G; 117 T; 0 U; 0 Other;

  Query Match   8.9%; Score 185.4; DB 6; Length 522;
  Best Local Similarity   73.8%; Pred. No. 1.9e-34;
  Matches 401; Conservative   0; Mismatches   91; Indels   51; Gaps   11;

```
Qy       1532 GCTAAGGTCTTACTCCTGGAAATCCCAACACCAGTTCTTCAGGGACTGCTGTGAGGCAGT 1591
              | |||||||||    ||||||||| || ||||| ||||||| |||||  || |||||||
Db        519 GTTAAGGTCTTCAGTCTGGAAATCACAGCACCA-TTCTTCAGAGACTGTGGTCAGGCAGT 461

Qy       1592 GCCTTATGCAGGTCTTGTCCTTGGCCATCACTGTCTGGTTCCCAGCCCAGCAC-ATGTGA 1650
              ||||||||||||||| |||||||             ||| ||||||||||||  |||||
Db        460 GCCTTATGCAGGTCTGGTCCTTG------------GGTCCCCAGCCCAGCACTGTGTGA 414

Qy       1651 CATGAGGACATGACATGCCCGAACCACCCAGCACCACATGCTCCATGTCAAGTGTGTACG 1710
              || ||||||| ||||| ||    | || ||||||||||||||| ||||||||| || | ||
Db        413 CACGAGGACACGACATCCCAACAACAACCAGCACCACATGCTACATGTCAAGAGTATGCG 354

Qy       1711 TGGAGACCACTGGCTCCCAGGC----CTGTGCTCAGAGAGGGTGTGCAGTCCTACGTGTG 1766
              |||||||  | || | || |||      |||||| ||| |||||||||| ||  | ||||
Db        353 TGGAGACTGCAGGGTGCCGGGCATGTGTGTGCTTAGAAAGGGTGTGCAGCCCACCATGTG 294

Qy       1767 CTGGGGGGGACGACGGTGACCTGTGCTTGCTTGCTTTTAAAATGGTGCTTGGACGTTTTA 1826
              ||      | | |  ||||||||||||||||| ||||||||||||||||| || |||||
Db        293 CT------CAATAGGAGTACCTGTGCTTGCTTGC-TTTAAAATGGTGCTTGAACATTTTA 241

Qy       1827 AGGTTAAAAACAATCCGACTCCATATGATTTAGGGCTCCTCCACCCTGGGGTGGCCCCTA 1886
              |||||||||||||||| || ||  ||  | |||| ||||||||| |||||||||||||| |
Db        240 AGGTTAAAAACAATCTGATTCTGTA--ACTTAGTGCTCCTCCATCCTGGGGTGGCCCC-A 184

Qy       1887 TGCTGTCTGCTTGGATCTCAAAGTCTTGGT---------ACTCGGCACTGTCAGACTCCA 1937
              |||||||||||||| |||||  |||  | |          ||  | ||    |  |||
Db        183 TGCTGTCTGCTTGGGTCTCATGGTCCCGATAACCAGATCACATGTCAGACCCCACCCCCA 124

Qy       1938 CCCCATGTATCCTTTTTGTTTCTCTTGTGCTTTTTTTGGACTTCCCAACCTGAGCCTA-- 1995
              |||||      | ||| | | |||  ||    || |||||||| |||| ||||||||||
Db        123 CCCCACCATGCTTTTGTTTCTCTTGTGCTGTTCTTTTGGACGTCCCCACCTGAGCCTAAG 64

Qy       1996 AGGTTTTATTTTATATGTG-----------CTTCAATATCAACAATGTAAACCTCACTTT 2044
              |||||||||||| ||||||          ||||| ||||||||||||||||||||||||
Db         63 AGGTTTTATTTTGTATGTGTTATGACTGTTCTTCACTATCAACAATGTAAACCTCACTTT 4

Qy       2045 ATT 2047
              |||
Db          3 ATT 1
```

RESULT 15
ACH24999
ID ACH24999 standard; cDNA; 477 BP.
XX
AC ACH24999;
XX
DT 13-OCT-2003 (first entry)
XX
DE Human adult ovary cDNA #3379.
XX
KW Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
KW genome mapping; biodiversity; genetic disorder.
XX
OS Homo sapiens.

XX
PN US2003073623-A1.
XX
PD 17-APR-2003.
XX
PF 30-JUL-2001; 2001US-00918995.
XX
PR 30-JUL-2001; 2001US-00918995.
XX
PA (DRMA/) DRMANAC R T.
PA (LABA/) LABAT I.
PA (STAC/) STACHE-CRAIN B.
PA (DICK/) DICKSON M C.
PA (JONE/) JONES L W.
XX
PI Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX
DR WPI; 2003-615964/58.
XX
PT New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridization probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.
XX
PS Claim 1; SEQ ID NO 12211; 44pp; English.
XX
CC The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversities, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030073623
XX
SQ Sequence 477 BP; 120 A; 78 C; 86 G; 183 T; 0 U; 10 Other;

Query Match 4.7%; Score 98.2; DB 8; Length 477;
Best Local Similarity 78.9%; Pred. No. 1.9e-13;
Matches 157; Conservative 0; Mismatches 33; Indels 9; Gaps 3;

```
Qy       1103 ACTTTCTAAAATTAAGTTTTATATGTTTTTGGGCAATATTTTGTCTTAAGATATATTTTT 1162
             | | | || |||| ||||||| ||| ||| ||||| ||||||||||||||||||||||||
Db         61 AATCTTTACAATTCAGTTTTACATG-TTTCGGGCAGTATTTTGTCTTAAGATATATTTTT 119

Qy       1163 TAAACTTTTTATACTTTA-------GATTTTTTTCAGCTATTTTCTTAAAAGTATATTTT 1215
             |||||||||||||| |||         ||||||||||||||||||||||||||||||||
Db        120 TAAACTTTTTATACCTTATCGGTTGAGATTTTTTCAGCTATTTTCTTAAAAGTATATTTT 179

Qy       1216 TTCTACAAACATCCTCTGCTGCTACATTAGAAACATTTATAACCTAAATACGATTGGTGT 1275
             ||||| ||||||||| |||||||||||||| ||| |||||| |||||| |      ||
Db        180 TTCTATAAACATCCTTTGCTGCTACATTAG-AACTTTTATAGCCTAAACAATTGCAGTTG 238

Qy       1276 GTCATTTTAAAGGTTTAAA 1294
             ||   ||| |   ||| ||
Db        239 GTGTGTTTCATTTTTTAA 257
```

Search completed: September 10, 2004, 11:11:26
Job time : 769 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: September 10, 2004, 10:58:30 ; Search time 207 Seconds
(without alignments)
5581.680 Million cell updates/sec

Title: US-10-076-069-1
Perfect score: 2082
Sequence: 1 ccccgcgtcggtcttccacc.....atggaaaaaaaaaaaaaaaa 2082

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	425.2	20.4	2629	4	US-09-620-312D-773	Sequence 773, App
	2	190	9.1	396	4	US-09-640-173-167	Sequence 167, App
	3	190	9.1	396	4	US-09-713-550-167	Sequence 167, App
c	4	47	2.3	1172	1	US-07-945-288-9	Sequence 9, Appli
c	5	47	2.3	1172	1	US-08-462-831-9	Sequence 9, Appli
c	6	47	2.3	1172	1	US-08-461-809-9	Sequence 9, Appli
c	7	47	2.3	1172	1	US-08-461-441-9	Sequence 9, Appli
c	8	47	2.3	1172	5	PCT-US93-08518-9	Sequence 9, Appli
	9	47	2.3	6801	4	US-10-204-708-62	Sequence 62, Appl
c	10	47	2.3	19124	2	US-08-487-826B-13	Sequence 13, Appl
	11	47	2.3	640681	4	US-09-790-988-1	Sequence 1, Appli
	12	46.6	2.2	10467	4	US-10-204-708-2	Sequence 2, Appli
	13	46	2.2	665	2	US-08-883-795A-36	Sequence 36, Appl
	14	45.4	2.2	11049	4	US-10-204-708-23	Sequence 23, Appl
	15	45.2	2.2	4140	3	US-08-894-731-2	Sequence 2, Appli
	16	45.2	2.2	6306	4	US-10-204-708-50	Sequence 50, Appl
c	17	45.2	2.2	640681	4	US-09-790-988-1	Sequence 1, Appli
	18	45	2.2	6040	4	US-10-204-708-70	Sequence 70, Appl
	19	45	2.2	10619	4	US-10-204-708-3	Sequence 3, Appli
	20	44.8	2.2	658	3	US-08-998-416-595	Sequence 595, App
c	21	44.6	2.1	2251	3	US-08-991-677-11	Sequence 11, Appl
c	22	44.4	2.1	152331	3	US-09-128-155-16	Sequence 16, Appl
c	23	44.2	2.1	2447	2	US-09-014-969-14	Sequence 14, Appl
	24	44.2	2.1	5666	4	US-10-204-708-29	Sequence 29, Appl
	25	44.2	2.1	11131	4	US-10-204-708-28	Sequence 28, Appl
c	26	43.8	2.1	1447	4	US-09-443-041A-27	Sequence 27, Appl
	27	43.8	2.1	5360	4	US-10-204-708-65	Sequence 65, Appl
c	28	43.6	2.1	2246	4	US-09-363-708-3	Sequence 3, Appli
c	29	43.6	2.1	2246	4	US-09-083-587-3	Sequence 3, Appli
	30	43.6	2.1	7304	4	US-10-204-708-44	Sequence 44, Appl
	31	43.6	2.1	8607	4	US-10-204-708-72	Sequence 72, Appl
c	32	43.4	2.1	6243	2	US-09-056-075-1	Sequence 1, Appli
c	33	43.2	2.1	2394	4	US-09-800-729-33	Sequence 33, Appl
c	34	43.2	2.1	5666	4	US-10-204-708-30	Sequence 30, Appl
	35	43.2	2.1	8961	4	US-10-204-708-80	Sequence 80, Appl
	36	43	2.1	396	4	US-09-640-173-33	Sequence 33, Appl
	37	43	2.1	396	4	US-09-713-550-33	Sequence 33, Appl
	38	43	2.1	6866	4	US-10-204-708-20	Sequence 20, Appl
	39	42.8	2.1	5219	4	US-10-204-708-51	Sequence 51, Appl
	40	42.8	2.1	11049	4	US-10-204-708-22	Sequence 22, Appl
	41	42.8	2.1	99916	4	US-09-816-095-3	Sequence 3, Appli
c	42	42.6	2.0	1390	4	US-09-704-139-1	Sequence 1, Appli
	43	42.6	2.0	5562	4	US-10-204-708-63	Sequence 63, Appl
c	44	42.6	2.0	6409	4	US-09-967-908A-1	Sequence 1, Appli
c	45	42.4	2.0	4254	2	US-08-443-639-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-09-620-312D-773
; Sequence 773, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 773
; LENGTH: 2629
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (73)..(798)
US-09-620-312D-773

Query Match 20.4%; Score 425.2; DB 4; Length 2629;
Best Local Similarity 74.1%; Pred. No. 1.4e-109;
Matches 557; Conservative 0; Mismatches 183; Indels 12; Gaps 1;

```
Qy        175 CGCCCGCCTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGA 234
              ||| | | |||||  ||||||||  |||| ||||||||||||| || || ||| || ||
Db         57 CGCGGGTCAGGACACAATGTTTGCACGAGGACTGAAGAGGAAATGTGTTGGCCACGAGGA 116
```

```
Qy        235 AGGAGTAGAGGGT------------TTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACA 282
Db        117 AGACGTGGAGGGAGCCCTGGCCGGCTTGAAGACAGTGTCCTCATACAGCCTGCAGCGGCA 176

Qy        283 GTCACTCCTGGACATGTCCCTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAA 342
Db        177 GTCGCTCCTGGACATGTCTCTGGTGAAGTTGCAGCTTTGCCACATGCTTGTGGAGCCCAA 236

Qy        343 TCTCTGCCGCTCGGTCCTCATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCA 402
Db        237 CCTGTGCCGCTCAGTCCTCATTGCCAACACGGTCCGGCAGATCCAAGAGGAGATGACGCA 296

Qy        403 GGATGGTGTGTGGCATGGGATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCT 462
Db        297 GGATGGGACGTGGCGCACAGTGGCACCCCAGGCTGCAGAGCGGGCGCCGCTCGACCGCTT 356

Qy        463 GGTGTCCACAGAGATCCTGTGTCGTACAGTGAGGGGAGCTGAGGAAGAGCACCCTGCTCC 522
Db        357 GGTCTCCACGGAGATCCTGTGCCGTGCAGCGTGGGGGCAAGAGGGGGCACATCCTGCTTC 416

Qy        523 TGAACTGGAAGATGCTCCCTTGCAAAACTCGGTTTCCGAGCTCCCCATCGTTGGCTCAGC 582
Db        417 TGGCTTGGGGGACGGCCACACACAGGGTCCAGTTTCTGACCTTTGCCCAGTCACCTCAGC 476

Qy        583 ACCAGGGCAAAGGAACCCTCAGAGCAGCCTCTGGGAGATGGACAGCCCACAAGAAAACAG 642
Db        477 ACAGGCACCAAGGCACCTGCAGAGCAGCGCCTGGGAGATGGATGGCCCTCGAGAAAACAG 536

Qy        643 GGGAAGCTTTCAGAAGTCACTGGACCAGATATTTGAGACCCTGGAGAACAAAAACTCCAG 702
Db        537 AGGAAGCTTTCACAAGTCACTTGATCAGATATTTGAAACGCTGGAGACTAAAAACCCCAG 596

Qy        703 TTCAGTGGAGGAACTCTTCTCAGATGTGGACAGCTCCTACTATGACCTGGACACAGTGCT 762
Db        597 CTGCATGGAAGAGCTGTTCTCAGACGTGGACAGCCCCTACTACGACCTGGACACAGTACT 656

Qy        763 AACAGGAATGATGAGTGGGACCAAGTCCAGTCTCTGCAATGGCCTTGAGGGCTTTGCTGC 822
Db        657 GACAGGCATGATGGGGGGTGCCAGGCCGGGCCCCTGCGAAGGGCTCGAGGGCTTGGCTCC 716

Qy        823 AGCCACCCCTCCTCCCAGTTCCACTTGCAAGTCTGACCTGGCTGAGCTGGACCATGTGGT 882
Db        717 GGCCACCCCAGGCCCTAGCTCCAGCTGCAAGTCCGACCTGGGCGAGCTGGACCACGTGGT 776

Qy        883 AGAGATTCTGGTGGAGACCTGAGAGGCCACCC 914
Db        777 GGAGATCCTGGTGGAGACCTGAGCAGGAGCCC 808
```

RESULT 2
US-09-640-173-167
; Sequence 167, Application US/09640173
; Patent No. 6613515
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: OVARIAN TUMOR SEQUENCES AND
; TITLE OF INVENTION: METHODS OF USE THEREFOR
; FILE REFERENCE: 210121.484C2
; CURRENT APPLICATION NUMBER: US/09/640,173
; CURRENT FILING DATE: 2000-08-15
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 167
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-640-173-167

Query Match 9.1%; Score 190; DB 4; Length 396;
Best Local Similarity 74.7%; Pred. No. 1e-43;
Matches 257; Conservative 0; Mismatches 75; Indels 12; Gaps 1;

```
Qy        175 CGCCCGCCTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGA 234
Db         52 CGCGGGTCAGGACACAATGTTTGCACGAGGACTGAAGAGGAAATGTGTTGGCCACGAGGA 111

Qy        235 AGGAGTAGAGGGT------------TTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACA 282
Db        112 AGACGTGGAGGGAGCCCTGGCCGGCTTGAAGACAGTGTCCTCATACAGCCTGCAGCGGCA 171

Qy        283 GTCACTCCTGGACATGTCCCTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAA 342
Db        172 GTCGCTCCTGGACATGTCTCTGGTGAAGTTGCAGCTTTGCCACATGCTTGTGGAGCCCAA 231

Qy        343 TCTCTGCCGCTCGGTCCTCATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCA 402
Db        232 TCTGTGCCGCTCAGTCCTCATTGCCAACACGGTCCGGCAGATCCAAGAGGAGATGACGCA 291

Qy        403 GGATGGTGTGTGGCATGGGATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCT 462
Db        292 GGATGGGACGTGGCGCACAGTGGCACCCCAGGCTGCAGAGCGGGCGCCGCTCGACCGCTT 351

Qy        463 GGTGTCCACAGAGATCCTGTGTCGTACAGTGAGGGGAGCTGAGG 506
Db        352 GGTCTCCACGGAGATCCTGTGCCGTGCAGCGTGGGGGCAAGAGG 395
```

RESULT 3
US-09-713-550-167
; Sequence 167, Application US/09713550
; Patent No. 6617109
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.484C4
; CURRENT APPLICATION NUMBER: US/09/713,550
; CURRENT FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 205
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 167
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-713-550-167

Query Match 9.1%; Score 190; DB 4; Length 396;
Best Local Similarity 74.7%; Pred. No. 1e-43;
Matches 257; Conservative 0; Mismatches 75; Indels 12; Gaps 1;

```
Qy        175 CGCCCGCCTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGA 234
Db         52 CGCGGGTCAGGACACAATGTTTGCACGAGGACTGAAGAGGAAATGTGTTGGCCACGAGGA 111

Qy        235 AGGAGTAGAGGGT------------TTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACA 282
Db        112 AGACGTGGAGGGAGCCCTGGCCGGCTTGAAGACAGTGTCCTCATACAGCCTGCAGCGGCA 171

Qy        283 GTCACTCCTGGACATGTCCCTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAA 342
Db        172 GTCGCTCCTGGACATGTCTCTGGTGAAGTTGCAGCTTTGCCACATGCTTGTGGAGCCCAA 231

Qy        343 TCTCTGCCGCTCGGTCCTCATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCA 402
Db        232 TCTGTGCCGCTCAGTCCTCATTGCCAACACGGTCCGGCAGATCCAAGAGGAGATGACGCA 291

Qy        403 GGATGGTGTGTGGCATGGGATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCT 462
Db        292 GGATGGGACGTGGCGCACAGTGGCACCCCAGGCTGCAGAGCGGGCGCCGCTCGACCGCTT 351

Qy        463 GGTGTCCACAGAGATCCTGTGTCGTACAGTGAGGGGAGCTGAGG 506
Db        352 GGTCTCCACGGAGATCCTGTGCCGTGCAGCGTGGGGGCAAGAGG 395
```

```
RESULT 4
US-07-945-288-9/c
; Sequence 9, Application US/07945288
; Patent No. 5433948
;  GENERAL INFORMATION:
;    APPLICANT:  Thomas, Wayne R.
;    APPLICANT:  Chua, Kaw-Yan
;    TITLE OF INVENTION:  CLONING AND SEQUENCING OF ALLERGENS FROM
;    TITLE OF INVENTION:  DERMATOPHAGOIDES (HOUSE DUST MITES)
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 STATE STREET, SUITE 510
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII TEXT
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/945,288
;      FILING DATE:  19920910
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  580,655
;      FILING DATE:  11 SEPTEMBER 1990
;      APPLICATION NUMBER:  458,642
;      FILING DATE:  13 FEBRUARY 1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MANDRAGOURAS, AMY E.
;      REGISTRATION NUMBER:  P36,207
;      REFERENCE/DOCKET NUMBER:  IPC-010CC (IMI-024)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-7400
;      TELEFAX:  (617) 227-5941
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1172 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..738
US-07-945-288-9

  Query Match             2.3%;  Score 47;  DB 1;  Length 1172;
  Best Local Similarity   56.0%;  Pred. No. 0.005;
  Matches   89;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy       1119 TTTTATATGTTTTTGGGCAATATTTTGTCTTAAGATATATTTTTTAAACTTTTTATACTT 1178
             |||| | | |||||      | |||| | ||    | | ||||||    ||||| |  ||
Db       1169 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1110

Qy       1179 TAGATTTTTTTCAGCTATTTTCTTAAAAGTATATTTTTTCTACAAACATCCTCTGCTGCT 1238
              |   ||||||| |  | | |    | ||  |||   |||||||  |||  |    |     |
Db       1109 TTTTTTTTTTTAATTTTTGTGAATGAATTTATTCATTTTCTAAAAATTTTGATTTTAAAT 1050

Qy       1239 ACATTAGAAACATTTATAACCTAAATACGATTGGTGTGT 1277
              | |||||  ||| | |||    ||||    ||  ||| |
Db       1049 AAATTAGTGACAATCATATAAGAAATTGTCTTTTTGTTT 1011


RESULT 5
US-08-462-831-9/c
; Sequence 9, Application US/08462831
; Patent No. 5552142
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  T CELL EPITOPES OF THE MAJOR ALLERGENS FROM
;    TITLE OF INVENTION:  DERMATOPHAGOIDES
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 STATE STREET, SUITE 510
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII TEXT
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/462,831
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/945,288
;      FILING DATE:  10 SEPTEMBER 1992
;      APPLICATION NUMBER:  US 580,655
;      FILING DATE:  11 SEPTEMBER 1990
;      APPLICATION NUMBER:  US 458,642
;      FILING DATE:  13 FEBRUARY 1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MANDRAGOURAS, AMY E.
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  IPC-010CC (IMI-024)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-7400
;      TELEFAX:  (617) 227-5941
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1172 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..738
US-08-462-831-9

  Query Match             2.3%;  Score 47;  DB 1;  Length 1172;
  Best Local Similarity   56.0%;  Pred. No. 0.005;
  Matches   89;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy       1119 TTTTATATGTTTTTGGGCAATATTTTGTCTTAAGATATATTTTTTAAACTTTTTATACTT 1178
             |||| | | |||||      | |||| | ||    | | ||||||    ||||| |  ||
Db       1169 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1110

Qy       1179 TAGATTTTTTTCAGCTATTTTCTTAAAAGTATATTTTTTCTACAAACATCCTCTGCTGCT 1238
              |   ||||||| |  | | |    | ||  |||   |||||||  |||  |    |     |
Db       1109 TTTTTTTTTTTAATTTTTGTGAATGAATTTATTCATTTTCTAAAAATTTTGATTTTAAAT 1050

Qy       1239 ACATTAGAAACATTTATAACCTAAATACGATTGGTGTGT 1277
              | |||||  ||| | |||    ||||    ||  ||| |
Db       1049 AAATTAGTGACAATCATATAAGAAATTGTCTTTTTGTTT 1011


RESULT 6
US-08-461-809-9/c
; Sequence 9, Application US/08461809
; Patent No. 5770202
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  T CELL EPITOPES OF THE MAJOR ALLERGENS FROM
;    TITLE OF INVENTION:  DERMATOPHAGOIDES
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
```

```
;       ADDRESSEE:  LAHIVE & COCKFIELD
;       STREET:  60 STATE STREET, SUITE 510
;       CITY:  BOSTON
;       STATE:  MA
;       COUNTRY:  USA
;       ZIP:  02109
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  ASCII TEXT
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/461,809
;       FILING DATE:
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/945,288
;       FILING DATE:  10 SEPTEMBER 1992
;       APPLICATION NUMBER:  US 580,655
;       FILING DATE:  11 SEPTEMBER 1990
;       APPLICATION NUMBER:  US 458,642
;       FILING DATE:  13 FEBRUARY 1990
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  MANDRAGOURAS, AMY E.
;       REGISTRATION NUMBER:  36,207
;       REFERENCE/DOCKET NUMBER:  IPC-010CC (IMI-024)
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 227-7400
;       TELEFAX:  (617) 227-5941
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1172 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..738
US-08-461-809-9

  Query Match             2.3%;  Score 47;  DB 1;  Length 1172;
  Best Local Similarity   56.0%;  Pred. No. 0.005;
  Matches   89;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy       1119 TTTTATATGTTTTTGGGCAATATTTTGTCTTAAGATATATTTTTTAAACTTTTTATACTT 1178
             |||| | | |||||      | |||| | ||    | | ||||||    ||||| |  ||
Db       1169 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1110

Qy       1179 TAGATTTTTTTCAGCTATTTTCTTAAAAGTATATTTTTTCTACAAACATCCTCTGCTGCT 1238
             |   ||||||| |  | | |    | ||  |||   ||||||| |||  |     |      |
Db       1109 TTTTTTTTTTTAATTTTTGTGAATGAATTTATTCATTTTCTAAAAATTTTGATTTTAAAT 1050

Qy       1239 ACATTAGAAACATTTATAACCTAAATACGATTGGTGTGT 1277
              | |||||  ||| | |||     ||||    ||  ||| |
Db       1049 AAATTAGTGACAATCATATAAGAAATTGTCTTTTTGTTT 1011


RESULT 7
US-08-461-441-9/c
; Sequence 9, Application US/08461441
; Patent No. 5773002
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  T CELL EPITOPES OF THE MAJOR ALLERGENS FROM
;    TITLE OF INVENTION:  DERMATOPHAGOIDES
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 STATE STREET, SUITE 510
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII TEXT
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/461,441
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/945,288
;      FILING DATE:  10 SEPTEMBER 1992
;      APPLICATION NUMBER:  US 580,655
;      FILING DATE:  11 SEPTEMBER 1990
;      APPLICATION NUMBER:  US 458,642
;      FILING DATE:  13 FEBRUARY 1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MANDRAGOURAS, AMY E.
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  IPC-010CC (IMI-024)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-7400
;      TELEFAX:  (617) 227-5941
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1172 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..738
US-08-461-441-9

  Query Match             2.3%;  Score 47;  DB 1;  Length 1172;
  Best Local Similarity   56.0%;  Pred. No. 0.005;
  Matches   89;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy       1119 TTTTATATGTTTTTGGGCAATATTTTGTCTTAAGATATATTTTTTAAACTTTTTATACTT 1178
             |||| | | |||||      | |||| | ||    | | ||||||    ||||| |  ||
Db       1169 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1110

Qy       1179 TAGATTTTTTTCAGCTATTTTCTTAAAAGTATATTTTTTCTACAAACATCCTCTGCTGCT 1238
             |   ||||||| |  | | |    | ||  |||   ||||||| |||  |     |      |
Db       1109 TTTTTTTTTTTAATTTTTGTGAATGAATTTATTCATTTTCTAAAAATTTTGATTTTAAAT 1050

Qy       1239 ACATTAGAAACATTTATAACCTAAATACGATTGGTGTGT 1277
              | |||||  ||| | |||     ||||    ||  ||| |
Db       1049 AAATTAGTGACAATCATATAAGAAATTGTCTTTTTGTTT 1011


RESULT 8
PCT-US93-08518-9/c
; Sequence 9, Application PC/TUS9308518
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  T CELL EPITOPES OF THE MAJOR ALLERGENS FROM
;    TITLE OF INVENTION:  DERMATOPHAGOIDES
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 STATE STREET, SUITE 510
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII TEXT
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/08518
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/945,288
;      FILING DATE:  10 SEPTEMBER 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MANDRAGOURAS, AMY E.
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  IPC-010CC (IMI-024)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-7400
;      TELEFAX:  (617) 227-5941
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1172 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..738
PCT-US93-08518-9

  Query Match             2.3%;  Score 47;  DB 5;  Length 1172;
  Best Local Similarity   56.0%;  Pred. No. 0.005;
  Matches   89;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy       1119 TTTTATATGTTTTTGGGCAATATTTTGTCTTAAGATATATTTTTTAAACTTTTTATACTT 1178
              |||| | | |||||      | |||| | ||    | | ||||||    ||||| |  ||
Db       1169 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1110

Qy       1179 TAGATTTTTTTCAGCTATTTTCTTAAAAGTATATTTTTTCTACAAACATCCTCTGCTGCT 1238
              |   ||||||| |  | | |   | ||  |||   ||||||| |||  |    |    |
Db       1109 TTTTTTTTTTTAATTTTTGTGAATGAATTTATTCATTTTCTAAAAATTTTGATTTTAAAT 1050

Qy       1239 ACATTAGAAACATTTATAACCTAAATACGATTGGTGTGT 1277
              | |||||  ||| | |||    ||||    ||  ||| |
Db       1049 AAATTAGTGACAATCATATAAGAAATTGTCTTTTTGTTT 1011


RESULT 9
US-10-204-708-62
; Sequence 62, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;  TITLE OF INVENTION:  by Assessing DNA Methylation
;  FILE REFERENCE: 5013.1012
;  CURRENT APPLICATION NUMBER: US/10/204,708
;  CURRENT FILING DATE:  2003-05-06
;  PRIOR APPLICATION NUMBER: PCT/EP01/03971
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 62
;   LENGTH: 6801
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-62

  Query Match             2.3%;  Score 47;  DB 4;  Length 6801;
  Best Local Similarity   51.2%;  Pred. No. 0.016;
  Matches  110;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy       1103 ACTTTCTAAAATTAAGTTTTATATGTTTTTGGGCAATATTTTGTCTTAAGATATATTTTT 1162
              | ||| ||   ||| ||    | | ||||||     |   || | |||| | || |||||
Db       5570 ATTTTATATTTTTAGGTAAGTTTTTTTTTTGTAATTTTAATTTTGTTAATAAATTTTTTT 5629

Qy       1163 TAAACTTTTTATACTTTAGATTTTTTTCAGCTATTTTCTTAAAAGTATATTTTTTCTACA 1222
              |    |||| |  ||   ||| ||||    | |||| ||    || |  ||||| ||
Db       5630 TTTTAGTTTTTTTTTTATTATTATTTTTTTTTTTTTTTTTTTCGGTTTCGTTTTTTTATT 5689

Qy       1223 AACATCCTCTGCTGCTACATTAGAAACATTTATAACCTAAATACGATTGGTGTGTCATTT 1282
                |  | || |  |  |||  ||   ||| |         ||   | ||| |   ||||
Db       5690 TTTTTATTTTGGTTTTTTATTTTAATTTTTTTTTTTTTATTGTATATTCGGTTTTGTATTT 5749

Qy       1283 TAAAGGTTTAAATAGAAAACTTCTTTTGTTACTGA 1317
                |   |||   |  ||   |  |||| ||| | |
Db       5750 AATTTTTTTTTGTTAAAGGTTATTTTTATTATTAA 5784


RESULT 10
US-08-487-826B-13/c
; Sequence 13, Application US/08487826B
; Patent No. 5993827
;  GENERAL INFORMATION:
;    APPLICANT:  Sim, Kim L.
;    APPLICANT:  Chitnis, Chetan
;    APPLICANT:  Miller, Louis H.
;    APPLICANT:  Peterson, David S.
;    APPLICANT:  Su, Xin-zhaun
;    APPLICANT:  Wellems, Thomas E.
;    TITLE OF INVENTION:  BINDING DOMAINS FROM PLASMODIUM VIVAX
;    TITLE OF INVENTION:  AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe Martens Olson & Bear
;      STREET:  620 Newport Center Drive  16th Floor
;      CITY:  Newport Beach
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/487,826B
;      FILING DATE:  10-SEP-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  NIH121.001CP1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  19124 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-487-826B-13

  Query Match             2.3%;  Score 47;  DB 2;  Length 19124;
  Best Local Similarity   51.2%;  Pred. No. 0.031;
```

```
  Matches  110;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy       1080 TCTATTTTCAATTAGATAGGTGAACTTTCTAAAATTAAGTTTTATATGTTTTTGGGCAAT 1139
             | | |||     ||   | | | | |||| | |   ||   || |||      |||       |
Db      15611 TTTTTTTAATTTTTTTTTGATAATCTTTTTCATTTTTTATTCTATCAAAATTTATATTTT 15552

Qy       1140 ATTTTGTCTTAAGATATATTTTTTAAACTTTTTATACTTTAGATTTTTTTCAGCTATTTT 1199
             ||| |    ||    ||| ||||  ||| ||||| | ||||   |||||||     | || |
Db      15551 ATTATAATTTTTATTATTTTTTAAAAAATTTTTCTCCTTTTTTTTTTTTTTTTTTTTTAT 15492

Qy       1200 CTTAAAAGTATATTTTTTCTACAAACATCCTCTGCTGCTACATTAGAAACATTTATAACC 1259
              |||| |    | ||||||  ||       |  |||   |  |    ||  | | |||| |
Db      15491 TTTAATAAATTTTTTTTTTATATTTCATTTTTCTTTTTTTCATTTTAATAAATTTGTTTTT 15432

Qy       1260 TAAATACGATTGGTGTGTCATTTTAAAGGTTTAAA 1294
              | | |  ||    | | | |  | |  | ||||
Db      15431 ATATTTCTTTTTTAATATATAAATACATATATATAAA 15397


RESULT 11
US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
;  APPLICANT: SHIGENOBU, SHUJI
;  APPLICANT:  WATANABE, HIDEMI
;  APPLICANT:  HATTORI, MASAHIRA
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
;  FILE REFERENCE: 081356/0159
;  CURRENT APPLICATION NUMBER: US/09/790,988
;  CURRENT FILING DATE:  2001-02-23
;  PRIOR APPLICATION NUMBER: JP2000-107160
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 640681
;   TYPE: DNA
;   ORGANISM: Buchnera sp.
US-09-790-988-1

  Query Match             2.3%;  Score 47;  DB 4;  Length 640681;
  Best Local Similarity   50.7%;  Pred. No. 0.3;
  Matches  113;  Conservative    0;  Mismatches  110;  Indels    0;  Gaps    0;

Qy       1092 TAGATAGGTGAACTTTCTAAAATTAAGTTTTATATGTTTTTGGGCAATATTTTGTCTTAA 1151
             |||  |  |    | | |||||||||    || | |  ||||    ||| || | | ||
Db     325018 TAGTAATTTTTGATATTTAAAATTAATAATTTTTTTGTTTTTTAAAATTTTATTTTTTGT 325077

Qy       1152 GATATATTTTTTAAACTTTTTATACTTTAGATTTTTTTCAGCTATTTTCTTAAAAGTATA 1211
              || |   |   |    |||  |    || | |  |       |   ||   |  |   ||
Db     325078 AATTTTCATAGAACGTATTTCTTTAAATAAAGTAATAATTTTTTAATTTAAATTAAACTA 325137

Qy       1212 TTTTTTCTACAAACATCCTCTGCTGCTACATTAGAAACATTTATAACCTAAATACGATTG 1271
              ||||| | || |      | |  |    |  ||| ||| |||| | |  ||||||   |
Db     325138 ATTTTTATTCATATTATATTTTTTATAAATTTAAAAATATTTCTTAAATAAATATTTTAA 325197

Qy       1272 GTGTGTCATTTTAAAGGTTTAAATAGAAAACTTCTTTTGTTAC 1314
              |||||  ||  |||  | ||||||  ||   | || | ||||
Db     325198 ATGTGTATTTAAAAAAATATAAATAATAATTATTTTATTTTAC 325240


RESULT 12
US-10-204-708-2
; Sequence 2, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;  TITLE OF INVENTION:  by Assessing DNA Methylation
;  FILE REFERENCE: 5013.1012
;  CURRENT APPLICATION NUMBER: US/10/204,708
;  CURRENT FILING DATE:  2003-05-06
;  PRIOR APPLICATION NUMBER: PCT/EP01/03971
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 2
;   LENGTH: 10467
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-2

  Query Match             2.2%;  Score 46.6;  DB 4;  Length 10467;
  Best Local Similarity   49.4%;  Pred. No. 0.027;
  Matches  121;  Conservative    0;  Mismatches  124;  Indels    0;  Gaps    0;

Qy        972 GAGACAGATAAGCACTTGTCCTAAGAGGGGCTCTGGCTCTTGAGCTCATTATCCTTTTGT 1031
              ||||  |||    | ||||  ||| ||||  | |      ||    |  ||    |||  |
Db       1848 GAGAAGGATTTTGATTTGTGTTAATAGGGTTTATAAAAATTTTTTTTTTTGAATTTTATT 1907

Qy       1032 GTGACATTGGACTCACTGTGGAGGATGGTGTGTCACAGCTATGTCTAGTCTATTTTCAAT 1091
                |   || |        ||||   || |||| | | |  ||||| |    |  ||
Db       1908 TGGGTTTTCGTTGTGTGGTGGTTAATTGTGTTTAATAAATATGTTTTAGTTTATTGTGGA 1967

Qy       1092 TAGATAGGTGAACTTTCTAAAATTAAGTTTTATATGTTTTTGGGCAATATTTTGTCTTAA 1151
              | ||    |    |||  |||| |   |||| || | ||| | |  ||| |    ||
Db       1968 TTGAGTATTTTTTTTTTAAAAAATTTTTTTTTTAGGGTTTGGTGATATAGTAAAGGTTTT 2027

Qy       1152 GATATATTTTTTAAACTTTTTATACTTTAGATTTTTTTCAGCTATTTTCTTAAAAGTATA 1211
                | ||||||    ||||  |  |||   || ||||     ||||| | | |  |||
Db       2028 TTATTTTTTTTTTTTTTTTTGTTTTTTTTTTTTATTTTTTATAATTTTTTAATAGTTATT 2087

Qy       1212 TTTTT 1216
              | ||
Db       2088 GTATT 2092


RESULT 13
US-08-883-795A-36
; Sequence 36, Application US/08883795A
; Patent No. 5985607
;  GENERAL INFORMATION:
;    APPLICANT:  Delcuve, Genevieve
;    APPLICANT:  Awang, Gregor
;    TITLE OF INVENTION:  Recombinant DNA Molecules and Expression
;    TITLE OF INVENTION:  Vectors for Tissue Plasminogen Activator
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BERESKIN & PARR
;      STREET:  40 King Street West
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5H 3Y2
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/883,795A
;      FILING DATE:  27-JUN-1997
;      CLASSIFICATION:  435
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gravelle, Micheline
;      REGISTRATION NUMBER:  40,261
;      REFERENCE/DOCKET NUMBER:  7841-062
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (416) 364-7311
;      TELEFAX:  (416) 361-1398
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  665 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      CLONE:  Rh 32
US-08-883-795A-36

  Query Match             2.2%;  Score 46;  DB 2;  Length 665;
  Best Local Similarity   52.0%;  Pred. No. 0.0067;
  Matches  129;  Conservative    0;  Mismatches  115;  Indels    4;  Gaps    1;

Qy       1066 ACAGCTATGTCTAGTCTATTTTCAATTAGATAGGTGAACTTTCTAAAATTAAGTTTTATA 1125
             | |  |||   ||   |||||  ||||| |||  | |   ||  || ||| |   ||| |
Db         50 ATATTTATAATTAAAATATTTATAATTAAATATTTTATAATTAAAATATTTATAATTAAA 109

Qy       1126 TGTTTTTGGGCAATATTTTGTCTTAAGATATATTTT----TTAAACTTTTTATACTTTAG 1181
             | ||||     | | | | | | |  | |||||||    |||||  | ||||||  || |
Db        110 TATTTTATAATTAAAATATTTATAATTAAATATTTTATAATTAAAATATTTATAATTAAA 169

Qy       1182 ATTTTTTTCAGCTATTTTCTTAAAAGTATATTTTTTCTACAAACATCCTCTGCTGCTACA 1241
               ||||  | |   |  |  |||  || || || |||| ||   | |   | |     || |
Db        170 TATTTTATAATTAAAATATTTATAATTAAATATTTTATAATTAAAATATTTATAATTAAA 229

Qy       1242 TTAGAAACATTTATAACCTAAATACGATTGGTGTGTCATTTTAAAGGTTTAAATAGAAAA 1301
              |      | | |||  ||  |  |||          | |    |||||   || |  |  |||
Db        230 TATTTTATAATTAAAATATTTATAATTAAATATTTTATAATTAAAATGTTTATAATTAAA 289

Qy       1302 CTTCTTTT 1309
                | ||| |
Db        290 TATTTTAT 297


RESULT 14
US-10-204-708-23
; Sequence 23, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;  TITLE OF INVENTION:  by Assessing DNA Methylation
;  FILE REFERENCE: 5013.1012
;  CURRENT APPLICATION NUMBER: US/10/204,708
;  CURRENT FILING DATE:  2003-05-06
;  PRIOR APPLICATION NUMBER: PCT/EP01/03971
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 23
;   LENGTH: 11049
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-23

  Query Match             2.2%;  Score 45.4;  DB 4;  Length 11049;
  Best Local Similarity   52.3%;  Pred. No. 0.061;
  Matches  124;  Conservative    0;  Mismatches  111;  Indels    2;  Gaps    1;

Qy       1112 AATTAAGTTTTATATGTTTTTGGGCAATATTTTGTCTTAAGATATATTTTTTAAACTTTT 1171
              | |||  |||| | | ||| | ||   | |||| | ||    |||  ||   |  ||||
Db       8235 ATTTAGTTTTTTTTTATTTATTGGACGTTTTTTTTTTTTTTTTATTATTAGAAGTTTTTT 8294

Qy       1172 TATACTTTAGATTTTTTTCAGCTATTTTCTTAAAAGTATATTTTTTCTACAAACATCCTC 1231
              | | |||   | ||||| |  |  ||| |  |   ||  || || |       ||  |
Db       8295 TTAAATTTTTTTATTTTTAATTTTGTTTTTATATTATAATATTATTGTTTTTTGATATTA 8354

Qy       1232 TGCTGCTACATTAGAAACAT--TTATAACCTAAATACGATTGGTGTGTCATTTTAAAGGT 1289
              |  |     | |||||| ||  ||||    ||  |  | |||   ||  |||||      |
Db       8355 TTTTATAGAAATAGAAAAATGGTTATTTAATATTTTTGGTTGAATTGGTATTTTTTTTTT 8414

Qy       1290 TTAAATAGAAAACTTCTTTTGTTACTGAGTCTCTACACTCCCAAGGCAACTGTAAAT 1346
              ||||| |||||  ||  | |||   | | |   | | |   ||   || ||| |||
Db       8415 TTAAAAAGAAATTTTTATGTGTGTATAAATAAGAAAAATGAAAATAGAATTGTTAAT 8471


RESULT 15
US-08-894-731-2
; Sequence 2, Application US/08894731
; Patent No. 6084089
; GENERAL INFORMATION:
;  APPLICANT: MINE, Toshiki
;  APPLICANT: OHYAMA, Akio
;  APPLICANT: HIYOSHI, Toru
;  APPLICANT: KASAOKA, Keisuke
;  TITLE OF INVENTION: COLD-INDUCIBLE PROMOTER SEQUENCE
;  FILE REFERENCE: 760-234P
;  CURRENT APPLICATION NUMBER: US/08/894,731
;  CURRENT FILING DATE: 1997-10-27
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 4140
;   TYPE: DNA
;   ORGANISM: Solanum tuberosum
US-08-894-731-2

  Query Match             2.2%;  Score 45.2;  DB 3;  Length 4140;
  Best Local Similarity   51.5%;  Pred. No. 0.037;
  Matches  104;  Conservative    0;  Mismatches   98;  Indels    0;  Gaps    0;

Qy       1122 TATATGTTTTTGGGCAATATTTTGTCTTAAGATATATTTTTTAAACTTTTTATACTTTAG 1181
              ||||| ||| |    || | || |   |||  || || || |     ||||   ||||
Db        237 TATATATTTATTAAAAAAAATTAGCTATAAATTAGATATTATTTTTATTTTTACCTTTTT 296

Qy       1182 ATTTTTTTCAGCTATTTTCTTAAAAGTATATTTTTTCTACAAACATCCTCTGCTGCTACA 1241
              | || |||       | || |||||| |||| ||  |  | ||  | |    |     || |
Db        297 AGTTATTTTCAAATTATTTTTAAAAATATAATTAATTCAAAATAAACTCATCAATATAGA 356

Qy       1242 TTAGAAACATTTATAACCTAAATACGATTGGTGTGTCATTTTAAAGGTTTAAATAGAAAA 1301
              | |  ||   |  | |  | | |||          |  | ||||    ||||||||||
Db        357 TAATTAAATCTCTTGAAATTATTACCTAGAAAAATATAAATGAAAGACCCAAATAGAAAA 416

Qy       1302 CTTCTTTTGTTACTGAGTCTCT 1323
                ||  |   | | | | |||||
Db        417 ATTAATAATTCAATAACTCTCT 438


Search completed: September 10, 2004, 14:20:52
Job time : 210 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: September 10, 2004, 12:42:41 ; Search time 699 Seconds
(without alignments)
14986.596 Million cell updates/sec

Title: US-10-076-069-1
Perfect score: 2082
Sequence: 1 ccccgcgtcggtcttccacc.....atggaaaaaaaaaaaaaaaa 2082

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3304383 seqs, 2515761380 residues

Total number of hits satisfying chosen parameters: 6608766

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2082	100.0	2082	14	US-10-076-069-1	Sequence 1, Appli
	2	428.4	20.6	2038	9	US-09-925-302-153	Sequence 153, App
	3	428.4	20.6	2038	13	US-09-925-302-153	Sequence 153, App
	4	428.4	20.6	2038	15	US-10-106-698-179	Sequence 179, App
	5	426.8	20.5	2082	14	US-10-076-069-3	Sequence 3, Appli
	6	425.2	20.4	2447	13	US-10-362-893-5	Sequence 5, Appli
	7	425.2	20.4	2629	15	US-10-037-270-773	Sequence 773, App
	8	425.2	20.4	2629	16	US-10-117-722-773	Sequence 773, App
	9	190	9.1	396	9	US-09-825-294-167	Sequence 167, App
	10	190	9.1	396	9	US-09-970-966-167	Sequence 167, App
	11	190	9.1	396	15	US-10-212-677-167	Sequence 167, App
	12	190	9.1	396	16	US-10-361-811-167	Sequence 167, App
	13	190	9.1	396	16	US-10-369-186-167	Sequence 167, App
c	14	185.4	8.9	522	9	US-09-917-800A-1285	Sequence 1285, Ap
	15	98.2	4.7	477	10	US-09-918-995-12211	Sequence 12211, A
	16	62.4	3.0	368	9	US-09-867-701-3214	Sequence 3214, Ap
	17	61.2	2.9	18988	13	US-10-221-714A-64	Sequence 64, Appl
	18	61.2	2.9	18988	15	US-10-311-455-674	Sequence 674, App
	19	61.2	2.9	18988	15	US-10-240-485-62	Sequence 62, Appl
	20	60.4	2.9	3673778	15	US-10-312-841-1	Sequence 1, Appli
	21	59	2.8	17721	15	US-10-311-455-1701	Sequence 1701, Ap
	22	56.2	2.7	9642	15	US-10-311-455-329	Sequence 329, App
c	23	55.8	2.7	960	15	US-10-198-846-6381	Sequence 6381, Ap
	24	55.2	2.7	5689	13	US-10-221-714A-148	Sequence 148, App
	25	55.2	2.7	5689	15	US-10-239-676-90	Sequence 90, Appl
	26	55.2	2.7	5689	15	US-10-240-453-100	Sequence 100, App
	27	55.2	2.7	83391	17	US-10-433-793-123	Sequence 123, App
	28	54.8	2.6	15548	15	US-10-311-455-2128	Sequence 2128, Ap
	29	54.6	2.6	5488	15	US-10-311-455-1429	Sequence 1429, Ap
	30	54.4	2.6	33053	17	US-10-433-793-36	Sequence 36, Appl
	31	54.4	2.6	37973	15	US-10-311-455-2169	Sequence 2169, Ap
	32	54	2.6	8136	15	US-10-311-455-528	Sequence 528, App
	33	54	2.6	8136	17	US-10-257-166-38	Sequence 38, Appl
	34	53.8	2.6	7589	15	US-10-240-453-264	Sequence 264, App
	35	53.6	2.6	113515	15	US-10-311-455-2147	Sequence 2147, Ap
	36	53.4	2.6	6128	15	US-10-311-455-29	Sequence 29, Appl
	37	52.6	2.5	6904	15	US-10-311-455-185	Sequence 185, App
	38	52.2	2.5	5273	15	US-10-311-455-848	Sequence 848, App
	39	52	2.5	15161	13	US-10-221-613-385	Sequence 385, App
	40	52	2.5	3673778	15	US-10-312-841-2	Sequence 2, Appli
	41	51.8	2.5	591	17	US-10-437-963-21828	Sequence 21828, A
	42	51.8	2.5	6154	13	US-10-221-613-70	Sequence 70, Appl
	43	51.8	2.5	17538	15	US-10-311-455-1130	Sequence 1130, Ap
	44	51.6	2.5	516	9	US-09-960-352-5785	Sequence 5785, Ap
	45	51.6	2.5	6478	15	US-10-239-676-124	Sequence 124, App

ALIGNMENTS

RESULT 1
US-10-076-069-1
; Sequence 1, Application US/10076069
; Publication No. US20020177214A1
; GENERAL INFORMATION:
; APPLICANT: JURECIC, ROLAND
; APPLICANT: NACHTMAN, RONALD
; TITLE OF INVENTION: HBPP, A NOVEL GENE WITH A ROLE IN HEMATOPOIETIC AND NEURAL DEVEL
; FILE REFERENCE: 39532-176599
; CURRENT APPLICATION NUMBER: US/10/076,069
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: US 60/268,923
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2082
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (191)..(901)
US-10-076-069-1

Query Match 100.0%; Score 2082; DB 14; Length 2082;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2082; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CCCCGCGTCGGTCTTCCACCTCACCTTTCGAGCTGGCCGCCGCTTGCTGTGCGCAGTTTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCCCGCGTCGGTCTTCCACCTCACCTTTCGAGCTGGCCGCCGCTTGCTGTGCGCAGTTTC 60

Qy         61 GGGGGACTGGACCTTCCCTGGCTTTTAGCAGCGCCGAGCGCCATGGCGACCCTTTGCTGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGGGGACTGGACCTTCCCTGGCTTTTAGCAGCGCCGAGCGCCATGGCGACCCTTTGCTGG 120

Qy        121 GCAGGTGACCGATTCCGGGTGCCCGAAGGAGCTGGCGTGGGTCTGCCTTGCAGCCGCCCG 180
```

```
Db       121 GCAGGTGACCGATTCCGGGTGCCCGAAGGAGCTGGCGTGGGTCTGCCTTGCAGCCGCCCG 180
Qy       181 CCTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGAAGGAGT 240
Db       181 CCTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGAAGGAGT 240
Qy       241 AGAGGGTTTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACAGTCACTCCTGGACATGTC 300
Db       241 AGAGGGTTTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACAGTCACTCCTGGACATGTC 300
Qy       301 CCTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAATCTCTGCCGCTCGGTCCT 360
Db       301 CCTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAATCTCTGCCGCTCGGTCCT 360
Qy       361 CATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCAGGATGGTGTGTGGCATGG 420
Db       361 CATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCAGGATGGTGTGTGGCATGG 420
Qy       421 GATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCTGGTGTCCACAGAGATCCT 480
Db       421 GATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCTGGTGTCCACAGAGATCCT 480
Qy       481 GTGTCGTACAGTGAGGGGAGCTGAGGAAGAGCACCCTGCTCCTGAACTGGAAGATGCTCC 540
Db       481 GTGTCGTACAGTGAGGGGAGCTGAGGAAGAGCACCCTGCTCCTGAACTGGAAGATGCTCC 540
Qy       541 CTTGCAAAACTCGGTTTCCGAGCTCCCCATCGTTGGCTCAGCACCAGGGCAAAGGAACCC 600
Db       541 CTTGCAAAACTCGGTTTCCGAGCTCCCCATCGTTGGCTCAGCACCAGGGCAAAGGAACCC 600
Qy       601 TCAGAGCAGCCTCTGGGAGATGGACAGCCCACAAGAAAACAGGGGAAGCTTTCAGAAGTC 660
Db       601 TCAGAGCAGCCTCTGGGAGATGGACAGCCCACAAGAAAACAGGGGAAGCTTTCAGAAGTC 660
Qy       661 ACTGGACCAGATATTTGAGACCCTGGAGAACAAAAACTCCAGTTCAGTGGAGGAACTCTT 720
Db       661 ACTGGACCAGATATTTGAGACCCTGGAGAACAAAAACTCCAGTTCAGTGGAGGAACTCTT 720
Qy       721 CTCAGATGTGGACAGCTCCTACTATGACCTGGACACAGTGCTAACAGGAATGATGAGTGG 780
Db       721 CTCAGATGTGGACAGCTCCTACTATGACCTGGACACAGTGCTAACAGGAATGATGAGTGG 780
Qy       781 GACCAAGTCCAGTCTCTGCAATGGCCTTGAGGGCTTTGCTGCAGCCACCCCTCCTCCCAG 840
Db       781 GACCAAGTCCAGTCTCTGCAATGGCCTTGAGGGCTTTGCTGCAGCCACCCCTCCTCCCAG 840
Qy       841 TTCCACTTGCAAGTCTGACCTGGCTGAGCTGGACCATGTGGTAGAGATTCTGGTGGAGAC 900
Db       841 TTCCACTTGCAAGTCTGACCTGGCTGAGCTGGACCATGTGGTAGAGATTCTGGTGGAGAC 900
Qy       901 CTGAGAGGCCACCCCAGTGGGCTAAGGGTGAGGCCACCAGTCCCCATGGAGCTCACGTGT 960
Db       901 CTGAGAGGCCACCCCAGTGGGCTAAGGGTGAGGCCACCAGTCCCCATGGAGCTCACGTGT 960
Qy       961 GTTGTGACCCAGAGACAGATAAGCACTTGTCCTAAGAGGGGCTCTGGCTCTTGAGCTCAT 1020
Db       961 GTTGTGACCCAGAGACAGATAAGCACTTGTCCTAAGAGGGGCTCTGGCTCTTGAGCTCAT 1020
Qy      1021 TATCCTTTTGTGTGACATTGGACTCACTGTGGAGGATGGTGTGTCACAGCTATGTCTAGT 1080
Db      1021 TATCCTTTTGTGTGACATTGGACTCACTGTGGAGGATGGTGTGTCACAGCTATGTCTAGT 1080
Qy      1081 CTATTTTCAATTAGATAGGTGAACTTTCTAAAATTAAGTTTTATATGTTTTTGGGCAATA 1140
Db      1081 CTATTTTCAATTAGATAGGTGAACTTTCTAAAATTAAGTTTTATATGTTTTTGGGCAATA 1140
Qy      1141 TTTTGTCTTAAGATATATTTTTTAAACTTTTTATACTTTAGATTTTTTTCAGCTATTTTC 1200
Db      1141 TTTTGTCTTAAGATATATTTTTTAAACTTTTTATACTTTAGATTTTTTTCAGCTATTTTC 1200
Qy      1201 TTAAAAGTATATTTTTTCTACAAACATCCTCTGCTGCTACATTAGAAACATTTATAACCT 1260
Db      1201 TTAAAAGTATATTTTTTCTACAAACATCCTCTGCTGCTACATTAGAAACATTTATAACCT 1260
Qy      1261 AAATACGATTGGTGTGTCATTTTAAAGGTTTAAATAGAAAACTTCTTTTGTTACTGAGTC 1320
Db      1261 AAATACGATTGGTGTGTCATTTTAAAGGTTTAAATAGAAAACTTCTTTTGTTACTGAGTC 1320
Qy      1321 TCTACACTCCCAAGGCAACTGTAAATGTAGCCGGCCGGGTGTTTACATGAGAGGCTCCAG 1380
Db      1321 TCTACACTCCCAAGGCAACTGTAAATGTAGCCGGCCGGGTGTTTACATGAGAGGCTCCAG 1380
Qy      1381 TATGGTCTACATTCTAGTAGAGCTTGAAAAGAACCATGCACAGCTCCACTGCCCCCTCAC 1440
Db      1381 TATGGTCTACATTCTAGTAGAGCTTGAAAAGAACCATGCACAGCTCCACTGCCCCCTCAC 1440
Qy      1441 TGGGTCTGCTCTGGCGGATCGGAGCTCTCTTCCTAGCCCCGTGTGCAGGATGGCTTTATT 1500
Db      1441 TGGGTCTGCTCTGGCGGATCGGAGCTCTCTTCCTAGCCCCGTGTGCAGGATGGCTTTATT 1500
Qy      1501 TATGCCTATTTATATGTAAATGCCACTGAAAGCTAAGGTCTTACTCCTGGAAATCCCAAC 1560
Db      1501 TATGCCTATTTATATGTAAATGCCACTGAAAGCTAAGGTCTTACTCCTGGAAATCCCAAC 1560
Qy      1561 ACCAGTTCTTCAGGGACTGCTGTGAGGCAGTGCCTTATGCAGGTCTTGTCCTTGGCCATC 1620
Db      1561 ACCAGTTCTTCAGGGACTGCTGTGAGGCAGTGCCTTATGCAGGTCTTGTCCTTGGCCATC 1620
Qy      1621 ACTGTCTGGTTCCCAGCCCAGCACATGTGACATGAGGACATGACATGCCCGAACCACCCA 1680
Db      1621 ACTGTCTGGTTCCCAGCCCAGCACATGTGACATGAGGACATGACATGCCCGAACCACCCA 1680
Qy      1681 GCACCACATGCTCCATGTCAAGTGTGTACGTGGAGACCACTGGCTCCCAGGCCTGTGCTC 1740
Db      1681 GCACCACATGCTCCATGTCAAGTGTGTACGTGGAGACCACTGGCTCCCAGGCCTGTGCTC 1740
Qy      1741 AGAGAGGGTGTGCAGTCCTACGTGTGCTGGGGGGGACGACGGTGACCTGTGCTTGCTTGC 1800
Db      1741 AGAGAGGGTGTGCAGTCCTACGTGTGCTGGGGGGGACGACGGTGACCTGTGCTTGCTTGC 1800
Qy      1801 TTTTAAAATGGTGCTTGGACGTTTTAAGGTTAAAAACAATCCGACTCCATATGATTTAGG 1860
Db      1801 TTTTAAAATGGTGCTTGGACGTTTTAAGGTTAAAAACAATCCGACTCCATATGATTTAGG 1860
Qy      1861 GCTCCTCCACCCTGGGGTGGCCCCTATGCTGTCTGCTTGGATCTCAAAGTCTTGGTACTC 1920
Db      1861 GCTCCTCCACCCTGGGGTGGCCCCTATGCTGTCTGCTTGGATCTCAAAGTCTTGGTACTC 1920
Qy      1921 GGCACTGTCAGACTCCACCCCATGTATCCTTTTTGTTTCTCTTGTGCTTTTTTTGGACTT 1980
Db      1921 GGCACTGTCAGACTCCACCCCATGTATCCTTTTTGTTTCTCTTGTGCTTTTTTTGGACTT 1980
Qy      1981 CCCAACCTGAGCCTAAGGTTTTATTTTATATGTGCTTCAATATCAACAATGTAAACCTCA 2040
Db      1981 CCCAACCTGAGCCTAAGGTTTTATTTTATATGTGCTTCAATATCAACAATGTAAACCTCA 2040
Qy      2041 CTTTATTAAAAGTATCCAGCAAATGGAAAAAAAAAAAAAAAA 2082
Db      2041 CTTTATTAAAAGTATCCAGCAAATGGAAAAAAAAAAAAAAAA 2082
```

RESULT 2
US-09-925-302-153
; Sequence 153, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12

```
;  NUMBER OF SEQ ID NOS: 896
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 153
;   LENGTH: 2038
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc feature
;   LOCATION: (1490)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (1508)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (1979)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (1992)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (2010)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-925-302-153

  Query Match             20.6%;  Score 428.4;  DB 9;  Length 2038;
  Best Local Similarity   74.3%;  Pred. No. 6.9e-98;
  Matches  559;  Conservative    0;  Mismatches  181;  Indels   12;  Gaps    1;

Qy        175 CGCCCGCCTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGA 234
              |||  | | |||||  ||||||||  |||| ||||||||||||| || || ||| || ||
Db         82 CGCGGGTCAGGACACAATGTTTGCACGAGGACTGAAGAGGAAATGTGTTGGCCACGAGGA 141

Qy        235 AGGAGTAGAGGGT------------TTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACA 282
              ||  || |||||              ||    || ||  | || || |||||||||| ||
Db        142 AGACGTGGAGGGAGCCCTGGCCGGCTTGAAGACAGTGTCCTCATACAGCCTGCAGCGGCA 201

Qy        283 GTCACTCCTGGACATGTCCCTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAA 342
              ||| |||||||||||||| || || ||| | ||||| || |||||||| ||||||||||
Db        202 GTCGCTCCTGGACATGTCTCTGGTGAAGTTGCAGCTTTGCCACATGCTTGTGGAGCCCAA 261

Qy        343 TCTCTGCCGCTCGGTCCTCATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCA 402
              ||| |||||||| |||||||| |||||||| |||||||||||||| ||||| ||||  ||
Db        262 TCTGTGCCGCTCAGTCCTCATTGCCAACACGGTCCGGCAGATCCAAGAGGAGATGACGCA 321

Qy        403 GGATGGTGTGTGGCATGGGATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCT 462
              ||||||   |||||      |||||||||||  || ||| ||||| ||  | || ||| |
Db        322 GGATGGGACGTGGCGCACAGTGGCACCCCAGGCTGCAGAGCGGGCGCCGYTCGACCGCTT 381

Qy        463 GGTGTCCACAGAGATCCTGTGTCGTACAGTGAGGGGAGCTGAGGAAGAGCACCCTGCTCC 522
              ||| ||||| |||||||||||| ||| ||| | ||||     |||| |  || |||||||
Db        382 GGTCTCCACGGAGATCCTGTGCCGTGCAGCGTGGGGGCAAGAGGGGGCACATCCTGCTCC 441

Qy        523 TGAACTGGAAGATGCTCCCTTGCAAAACTCGGTTTCCGAGCTCCCCATCGTTGGCTCAGC 582
              ||  |||  || |  | |    ||      | ||||| || ||    |    ||   ||||||
Db        442 TGGCTTGGGGGACGGCCACACACAGGGTCCAGTTTCTGACCTTTGCCCAGTCACCTCAGC 501

Qy        583 ACCAGGGCAAAGGAACCCTCAGAGCAGCCTCTGGGAGATGGACAGCCCACAAGAAAACAG 642
              ||  |  | |||| |||  |||||||||  ||||||||||||  |||| | |||||||||
Db        502 ACAGGCACCAAGGCACCTGCAGAGCAGCGCCTGGGAGATGGATGGCCCTCGAGAAAACAG 561

Qy        643 GGGAAGCTTTCAGAAGTCACTGGACCAGATATTTGAGACCCTGGAGAACAAAAACTCCAG 702
              ||||||||||| |||||||| || ||||||||||| || |||||||  |||||| ||||
Db        562 AGGAAGCTTTCACAAGTCACTTGATCAGATATTTGAAACGCTGGAGACTAAAAACCCCAG 621

Qy        703 TTCAGTGGAGGAACTCTTCTCAGATGTGGACAGCTCCTACTATGACCTGGACACAGTGCT 762
              |   |||| || |||||||| ||||||||| ||||||| |||||||||||||||| ||
Db        622 CTGCATGGAAGAGCTGTTCTCAGACGTGGACAGCCCCTACTACGACCTGGACACAGTACT 681

Qy        763 AACAGGAATGATGAGTGGGACCAAGTCCAGTCTCTGCAATGGCCTTGAGGGCTTTGCTGC 822
              ||||| |||||| | ||  ||| | |  | | |||| | || || |||||||| ||| |
Db        682 GACAGGCATGATGGGGGGTGCCAGGCCGGGCCCCTGCGAAGGGCTCGAGGGCTTGGCTCC 741

Qy        823 AGCCACCCCTCCTCCCAGTTCCACTTGCAAGTCTGACCTGGCTGAGCTGGACCATGTGGT 882
              ||||||||    || || |||| |||||||| |||||||  |||||||||| |||||
Db        742 GGCCACCCCRGGCCCTAGCTCCAGCTGCAAGTCCGACCTGGGCGAGCTGGACCACGTGGT 801

Qy        883 AGAGATTCTGGTGGAGACCTGAGAGGCCACCC 914
              ||||| ||||||||||||||||  |   |||
Db        802 GGAGATCCTGGTGGAGACCTGAGCAGGAGCCC 833


RESULT 3
US-09-925-302-153
; Sequence 153, Application US/09925302
; Publication No. US20030064072A9
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA104
;  CURRENT APPLICATION NUMBER: US/09/925,302
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05918
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 896
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 153
;   LENGTH: 2038
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc feature
;   LOCATION: (1490)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (1508)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (1979)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (1992)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (2010)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-925-302-153

  Query Match             20.6%;  Score 428.4;  DB 13;  Length 2038;
  Best Local Similarity   74.3%;  Pred. No. 6.9e-98;
  Matches  559;  Conservative    0;  Mismatches  181;  Indels   12;  Gaps    1;

Qy        175 CGCCCGCCTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGA 234
              |||  | | |||||  ||||||||  |||| ||||||||||||| || || ||| || ||
Db         82 CGCGGGTCAGGACACAATGTTTGCACGAGGACTGAAGAGGAAATGTGTTGGCCACGAGGA 141

Qy        235 AGGAGTAGAGGGT------------TTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACA 282
              ||  || |||||              ||    || ||  | || || |||||||||| ||
Db        142 AGACGTGGAGGGAGCCCTGGCCGGCTTGAAGACAGTGTCCTCATACAGCCTGCAGCGGCA 201

Qy        283 GTCACTCCTGGACATGTCCCTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAA 342
              ||| |||||||||||||| || || ||| | ||||| || |||||||| ||||||||||
Db        202 GTCGCTCCTGGACATGTCTCTGGTGAAGTTGCAGCTTTGCCACATGCTTGTGGAGCCCAA 261

Qy        343 TCTCTGCCGCTCGGTCCTCATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCA 402
              ||| |||||||| |||||||| |||||||| |||||||||||||| ||||| ||||  ||
Db        262 TCTGTGCCGCTCAGTCCTCATTGCCAACACGGTCCGGCAGATCCAAGAGGAGATGACGCA 321

Qy        403 GGATGGTGTGTGGCATGGGATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCT 462
              ||||||   |||||      |||||||||||  || ||| ||||| ||  | || ||| |
Db        322 GGATGGGACGTGGCGCACAGTGGCACCCCAGGCTGCAGAGCGGGCGCCGYTCGACCGCTT 381

Qy        463 GGTGTCCACAGAGATCCTGTGTCGTACAGTGAGGGGAGCTGAGGAAGAGCACCCTGCTCC 522
              ||| ||||| |||||||||||| ||| ||| | ||||     |||| |  || |||||||
Db        382 GGTCTCCACGGAGATCCTGTGCCGTGCAGCGTGGGGGCAAGAGGGGGCACATCCTGCTCC 441
```

```
Qy        523 TGAACTGGAAGATGCTCCCTTGCAAAACTCGGTTTCCGAGCTCCCCATCGTTGGCTCAGC 582
              ||   |||  || |  | |   ||      | ||||| || ||    |   ||   ||||||
Db        442 TGGCTTGGGGGACGGCCACACACAGGGTCCAGTTTCTGACCTTTGCCCAGTCACCTCAGC 501

Qy        583 ACCAGGGCAAAGGAACCCTCAGAGCAGCCTCTGGGAGATGGACAGCCCACAAGAAAACAG 642
              ||  |  | |||| |||  |||||||||  ||||||||||||  |||| | |||||||||
Db        502 ACAGGCACCAAGGCACCTGCAGAGCAGCGCCTGGGAGATGGATGGCCCTCGAGAAAACAG 561

Qy        643 GGGAAGCTTTCAGAAGTCACTGGACCAGATATTTGAGACCCTGGAGAACAAAAACTCCAG 702
              ||||||||||| ||||||||  || ||||||||||| || ||||||||  |||||| ||||
Db        562 AGGAAGCTTTCACAAGTCACTTGATCAGATATTTGAAACGCTGGAGACTAAAAACCCCAG 621

Qy        703 TTCAGTGGAGGAACTCTTCTCAGATGTGGACAGCTCCTACTATGACCTGGACACAGTGCT 762
              |   |||| || || |||||||| |||||||| ||||||| ||||||||||||||| ||
Db        622 CTGCATGGAAGAGCTGTTCTCAGACGTGGACAGCCCCTACTACGACCTGGACACAGTACT 681

Qy        763 AACAGGAATGATGAGTGGGACCAAGTCCAGTCTCTGCAATGGCCTTGAGGGCTTTGCTGC 822
              ||||| |||||| | ||  ||| | |  | |||| | || || |||||||| ||| |
Db        682 GACAGGCATGATGGGGGGTGCCAGGCCGGGCCCCTGCGAAGGGCTCGAGGGCTTGGCTCC 741

Qy        823 AGCCACCCCTCCTCCCAGTTCCACTTGCAAGTCTGACCTGGCTGAGCTGGACCATGTGGT 882
              ||||||||    || || ||||  |||||||| |||||||  ||||||||||| |||||
Db        742 GGCCACCCCRGGCCCTAGCTCCAGCTGCAAGTCCGACCTGGGCGAGCTGGACCACGTGGT 801

Qy        883 AGAGATTCTGGTGGAGACCTGAGAGGCCACCC 914
              ||||| ||||||||||||||||   |   |||
Db        802 GGAGATCCTGGTGGAGACCTGAGCAGGAGCCC 833
```

RESULT 4
US-10-106-698-179
; Sequence 179, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 179
; LENGTH: 2038
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1490)..(1490)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1508)..(1508)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1979)..(1979)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1992)..(1992)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (2010)..(2010)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-179

Query Match 20.6%; Score 428.4; DB 15; Length 2038;
Best Local Similarity 74.3%; Pred. No. 6.9e-98;
Matches 559; Conservative 0; Mismatches 181; Indels 12; Gaps 1;

```
Qy        175 CGCCCGCCTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGA 234
              |||  | | |||||  ||||||||  |||| ||||||||||||| || || ||| || ||
Db         82 CGCGGGTCAGGACACAATGTTTGCACGAGGACTGAAGAGGAAATGTGTTGGCCACGAGGA 141

Qy        235 AGGAGTAGAGGGT------------TTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACA 282
              ||  || |||||              ||    || ||  | || || ||||||||||| ||
Db        142 AGACGTGGAGGGAGCCCTGGCCGGCTTGAAGACAGTGTCCTCATACAGCCTGCAGCGGCA 201

Qy        283 GTCACTCCTGGACATGTCCCTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAA 342
              ||| |||||||||||||| || || ||| | |||||| || |||||||| ||||||||||
Db        202 GTCGCTCCTGGACATGTCTCTGGTGAAGTTGCAGCTTTGCCACATGCTTGTGGAGCCCAA 261

Qy        343 TCTCTGCCGCTCGGTCCTCATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCA 402
              ||| |||||||| |||||||| |||||||| |||||||||||||| ||||| ||||  ||
Db        262 TCTGTGCCGCTCAGTCCTCATTGCCAACACGGTCCGGCAGATCCAAGAGGAGATGACGCA 321

Qy        403 GGATGGTGTGTGGCATGGGATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCT 462
              ||||||   |||||       |||||||||||  || ||| ||||| ||  | || ||| |
Db        322 GGATGGGACGTGGCGCACAGTGGCACCCCAGGCTGCAGAGCGGGCGCCGYTCGACCGCTT 381

Qy        463 GGTGTCCACAGAGATCCTGTGTCGTACAGTGAGGGGAGCTGAGGAAGAGCACCCTGCTCC 522
              ||| ||||| |||||||||||| ||| ||| | |||     |||| |  || ||||||||
Db        382 GGTCTCCACGGAGATCCTGTGCCGTGCAGCGTGGGGGCAAGAGGGGGCACATCCTGCTCC 441

Qy        523 TGAACTGGAAGATGCTCCCTTGCAAAACTCGGTTTCCGAGCTCCCCATCGTTGGCTCAGC 582
              ||   |||  || |  | |   ||      | ||||| || ||    |   ||   ||||||
Db        442 TGGCTTGGGGGACGGCCACACACAGGGTCCAGTTTCTGACCTTTGCCCAGTCACCTCAGC 501

Qy        583 ACCAGGGCAAAGGAACCCTCAGAGCAGCCTCTGGGAGATGGACAGCCCACAAGAAAACAG 642
              ||  |  | |||| |||  |||||||||  ||||||||||||  |||| | |||||||||
Db        502 ACAGGCACCAAGGCACCTGCAGAGCAGCGCCTGGGAGATGGATGGCCCTCGAGAAAACAG 561

Qy        643 GGGAAGCTTTCAGAAGTCACTGGACCAGATATTTGAGACCCTGGAGAACAAAAACTCCAG 702
              |||||||||| |||||||| || |||||||||| || |||||||  |||||| ||||
Db        562 AGGAAGCTTTCACAAGTCACTTGATCAGATATTTGAAACGCTGGAGACTAAAAACCCCAG 621

Qy        703 TTCAGTGGAGGAACTCTTCTCAGATGTGGACAGCTCCTACTATGACCTGGACACAGTGCT 762
              |   |||| || || |||||||| |||||||| ||||||| |||||||||||||| ||
Db        622 CTGCATGGAAGAGCTGTTCTCAGACGTGGACAGCCCCTACTACGACCTGGACACAGTACT 681

Qy        763 AACAGGAATGATGAGTGGGACCAAGTCCAGTCTCTGCAATGGCCTTGAGGGCTTTGCTGC 822
              ||||| |||||| | ||  ||| | |  | |||| | || || |||||||| ||| |
Db        682 GACAGGCATGATGGGGGGTGCCAGGCCGGGCCCCTGCGAAGGGCTCGAGGGCTTGGCTCC 741

Qy        823 AGCCACCCCTCCTCCCAGTTCCACTTGCAAGTCTGACCTGGCTGAGCTGGACCATGTGGT 882
              ||||||||    || || ||||  |||||||| |||||||  ||||||||||| |||||
Db        742 GGCCACCCCRGGCCCTAGCTCCAGCTGCAAGTCCGACCTGGGCGAGCTGGACCACGTGGT 801

Qy        883 AGAGATTCTGGTGGAGACCTGAGAGGCCACCC 914
              ||||| ||||||||||||||||   |   |||
Db        802 GGAGATCCTGGTGGAGACCTGAGCAGGAGCCC 833
```

RESULT 5
US-10-076-069-3
; Sequence 3, Application US/10076069
; Publication No. US20020177214A1
; GENERAL INFORMATION:
; APPLICANT: JURECIC, ROLAND
; APPLICANT: NACHTMAN, RONALD
; TITLE OF INVENTION: HEPP, A NOVEL GENE WITH A ROLE IN HEMATOPOIETIC AND NEURAL DEVELO
; FILE REFERENCE: 39532-176599
; CURRENT APPLICATION NUMBER: US/10/076,069
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: US 60/268,923
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 2082
; TYPE: DNA
; ORGANISM: Homo sapiens

```
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (75)..(797)
US-10-076-069-3

  Query Match             20.5%;  Score 426.8;  DB 14;  Length 2082;
  Best Local Similarity   74.2%;  Pred. No. 1.8e-97;
  Matches  558;  Conservative    0;  Mismatches  182;  Indels   12;  Gaps    1;

Qy        175 CGCCCGCCTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGA 234
              |||  | | |||||  ||||||||  ||||| ||||||||||||| || || ||| || ||
Db         59 CGCGGGTCAGGACACAATGTTTGCACGAGGACTGAAGAGGAAATGTGTTGGCCACGAGGA 118

Qy        235 AGGAGTAGAGGGT------------TTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACA 282
              ||  || |||||             ||    || ||  | || || ||||||||||| ||
Db        119 AGACGTGGAGGGAGCCCTGGCCGGCTTGAAGACAGTGTCCTCATACAGCCTGCAGCGGCA 178

Qy        283 GTCACTCCTGGACATGTCCCTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAA 342
              ||| ||||||||||||||| || || ||| | ||||| || |||||||| ||||||||||
Db        179 GTCGCTCCTGGACATGTCTCTGGTGAAGTTGCAGCTTTGCCACATGCTTGTGGAGCCCAA 238

Qy        343 TCTCTGCCGCTCGGTCCTCATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCA 402
              || |||||||| |||||||| |||||||| |||||||||||||||| ||||| ||||  ||
Db        239 CCTGTGCCGCTCAGTCCTCATTGCCAACACGGTCCGGCAGATCCAAGAGGAGATGACGCA 298

Qy        403 GGATGGTGTGTGGCATGGGATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCT 462
              ||||||   |||||       |||||||||||  || ||| ||||| ||  | || ||| |
Db        299 GGATGGGACGTGGCGCACAGTGGCACCCCAGGCTGCAGAGCGGGCGCCGCTCGACCGCTT 358

Qy        463 GGTGTCCACAGAGATCCTGTGTCGTACAGTGAGGGGAGCTGAGGAAGAGCACCCTGCTCC 522
              ||| ||||| |||||||||||| ||| ||| | ||||    |||| |  || ||||||||
Db        359 GGTCTCCACGGAGATCCTGTGCCGTGCAGCGTGGGGGCAAGAGGGGGCACATCCTGCTCC 418

Qy        523 TGAACTGGAAGATGCTCCCTTGCAAAACTCGGTTTCCGAGCTCCCCATCGTTGGCTCAGC 582
              ||   |||  || |  | |   ||      | ||||| || ||   |   ||   ||||||
Db        419 TGGCTTGGGGGACGGCCACACACAGGGTCCAGTTTCTGACCTTTGCCCAGTCACCTCAGC 478

Qy        583 ACCAGGGCAAAGGAACCCTCAGAGCAGCCTCTGGGAGATGGACAGCCCACAAGAAAACAG 642
              ||  |  | |||| |||  |||||||||  ||||||||||||  |||| | |||||||||
Db        479 ACAGGCACCAAGGCACCTGCAGAGCAGCGCCTGGGAGATGGATGGCCCTCGAGAAAACAG 538

Qy        643 GGGAAGCTTTCAGAAGTCACTGGACCAGATATTTGAGACCCTGGAGAACAAAAACTCCAG 702
              ||||||||||| |||||||| || |||||||||| || |||||||  |||||| ||||
Db        539 AGGAAGCTTTCACAAGTCACTTGATCAGATATTTGAAACGCTGGAGACTAAAAACCCCAG 598

Qy        703 TTCAGTGGAGGAACTCTTCTCAGATGTGGACAGCTCCTACTATGACCTGGACACAGTGCT 762
              |   |||| || || |||||||| |||||||| ||||||| |||||||||||||||||| ||
Db        599 CTGCATGGAAGAGCTGTTCTCAGACGTGGACAGCCCCTACTACGACCTGGACACAGTACT 658

Qy        763 AACAGGAATGATGAGTGGGACCAAGTCCAGTCTCTGCAATGGCCTTGAGGGCTTTGCTGC 822
              |||||  |||||| | ||  ||| | |  | | |||| | || || |||||||| ||| |
Db        659 GACAGGCATGATGGGGGGTGCCAGGCCGGGCCCCTGCGAAGGGCTCGAGGGCTTGGCTCC 718

Qy        823 AGCCACCCCTCCTCCCAGTTCCACTTGCAAGTCTGACCTGGCTGAGCTGGACCATGTGGT 882
              ||||||||    || || ||||  |||||||| |||||||  |||||||||| |||||
Db        719 GGCCACCCCAGGCCCTAGCTCCAGCTGCAAGTCCGACCTGGGCGAGCTGGACCACGTGGT 778

Qy        883 AGAGATTCTGGTGGAGACCTGAGAGGCCACCC 914
              ||||| ||||||||||||||||  |   |||
Db        779 GGAGATCCTGGTGGAGACCTGAGCAGGAGCCC 810


RESULT 6
US-10-362-893-5
; Sequence 5, Application US/10362893
; Publication No. US20030211525A1
; GENERAL INFORMATION:
;  APPLICANT: Walker, Michael G.
;  APPLICANT:  Jung, Kenneth
;  TITLE OF INVENTION: GENES EXPRESSED IN THE CELL CYCLE
;  FILE REFERENCE: PB-0015 USN
;  CURRENT APPLICATION NUMBER: US/10/362,893
;  CURRENT FILING DATE:  2003-02-25
;  PRIOR APPLICATION NUMBER: US 60/229,253
;  PRIOR FILING DATE: 2000-08-30
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PERL Program
; SEQ ID NO 5
;   LENGTH: 2447
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: 225657.4
US-10-362-893-5

 .Query Match             20.4%;  Score 425.2;  DB 13;  Length 2447;
  Best Local Similarity   74.1%;  Pred. No. 5.1e-97;
  Matches  557;  Conservative    0;  Mismatches  183;  Indels   12;  Gaps    1;

Qy        175 CGCCCGCCTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGA 234
              |||  | | |||||  ||||||||  ||||| ||||||||||||| || || ||| || ||
Db         75 CGCGGGTCAGGACACAATGTTTGCACGAGGACTGAAGAGGAAATGTGTTGGCCACGAGGA 134

Qy        235 AGGAGTAGAGGGT------------TTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACA 282
              ||  || |||||             ||    || ||  | || || ||||||||||| ||
Db        135 AGACGTGGAGGGAGCCCTGGCCGGCTTGAAGACAGTGTCCTCATACAGCCTGCAGCGGCA 194

Qy        283 GTCACTCCTGGACATGTCCCTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAA 342
              ||| ||||||||||||||| || || ||| | ||||| || |||||||| ||||||||||
Db        195 GTCGCTCCTGGACATGTCTCTGGTGAAGTTGCAGCTTTGCCACATGCTTGTGGAGCCCAA 254

Qy        343 TCTCTGCCGCTCGGTCCTCATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCA 402
              || |||||||| |||||||| |||||||| |||||||||||||||| ||||| ||||  ||
Db        255 CCTGTGCCGCTCAGTCCTCATTGCCAACACGGTCCGGCAGATCCAAGAGGAGATGACGCA 314

Qy        403 GGATGGTGTGTGGCATGGGATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCT 462
              ||||||   |||||       |||||||||||  || ||| ||||| ||  | || ||| |
Db        315 GGATGGGACGTGGCGCACAGTGGCACCCCAGGCTGCAGAGCGGGCGCCGCTCGACCGCTT 374

Qy        463 GGTGTCCACAGAGATCCTGTGTCGTACAGTGAGGGGAGCTGAGGAAGAGCACCCTGCTCC 522
              ||| ||||| |||||||||||| ||| ||| | ||||    |||| |  || ||||||||
Db        375 GGTCTCCACGGAGATCCTGTGCCGTGCAGCGTGGGGGCAAGAGGGGGCACATCCTGCTCC 434

Qy        523 TGAACTGGAAGATGCTCCCTTGCAAAACTCGGTTTCCGAGCTCCCCATCGTTGGCTCAGC 582
              ||   |||  || |  | |   ||      | ||||| || ||   |   ||   ||||||
Db        435 TGGCTTGGGGGACGGCCACACACAGGGTCCAGTTTCTGACCTTTGCCCAGTCACCTCAGC 494

Qy        583 ACCAGGGCAAAGGAACCCTCAGAGCAGCCTCTGGGAGATGGACAGCCCACAAGAAAACAG 642
              ||  |  | |||| |||  |||||||||  ||||||||||||  |||| | |||||||||
Db        495 ACAGGCACCAAGGCACCTGCAGAGCAGCGCCTGGGAGATGGATGGCCCTCGAGAAAACAG 554

Qy        643 GGGAAGCTTTCAGAAGTCACTGGACCAGATATTTGAGACCCTGGAGAACAAAAACTCCAG 702
              ||||||||||| |||||||| || |||||||||| || |||||||  |||||| ||||
Db        555 AGGAAGCTTTCACAAGTCACTTGATCAGATATTTGAAACGCTGGAGACTAAAAACCCCAG 614

Qy        703 TTCAGTGGAGGAACTCTTCTCAGATGTGGACAGCTCCTACTATGACCTGGACACAGTGCT 762
              |   |||| || || |||||||| |||||||| ||||||| |||||||||||||||||| ||
Db        615 CTGCATGGAAGAGCTGTTCTCAGACGTGGACAGCCCCTACTACGACCTGGACACAGTACT 674

Qy        763 AACAGGAATGATGAGTGGGACCAAGTCCAGTCTCTGCAATGGCCTTGAGGGCTTTGCTGC 822
              |||||  |||||| | ||  ||| | |  | | |||| | || || |||||||| ||| |
Db        675 GACAGGCATGATGGGGGGTGCCAGGCCGGGCCCCTGCGAAGGGCTCGAGGGCTTGGCTCC 734

Qy        823 AGCCACCCCTCCTCCCAGTTCCACTTGCAAGTCTGACCTGGCTGAGCTGGACCATGTGGT 882
              ||||||||    || || ||||  |||||||| |||||||  |||||||||| |||  |
Db        735 GGCCACCCCAGGCCCTAGCTCCAGCTGCAAGTCCGACCTGGGCGAGCTGGACCACGTGAT 794

Qy        883 AGAGATTCTGGTGGAGACCTGAGAGGCCACCC 914
              ||||| ||||||||||||||||  |   |||
Db        795 GGAGATCCTGGTGGAGACCTGAGCAGGAGCCC 826


RESULT 7
US-10-037-270-773
; Sequence 773, Application US/10037270
```

```
; Publication No. US20030104529A1
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi, Vinod
;  APPLICANT:  Zhang, Jie
;  APPLICANT:  Ren, Feiyan
;  APPLICANT:  Chen, Rui-hong
;  APPLICANT:  Zhao, Qing A.
;  APPLICANT:  Wehrman, Tom
;  APPLICANT:  Xue, Aidong J.
;  APPLICANT:  Yang, Yonghong
;  APPLICANT:  Wang, Jian-Rui
;  APPLICANT:  Zhou, Ping
;  APPLICANT:  Ma, Yunqing
;  APPLICANT:  Wang, Dunrui
;  APPLICANT:  Wang, Zhiwei
;  APPLICANT:  Tillinghast, John
;  APPLICANT:  Drmanac, Radoje T.
;  TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and
;  TITLE OF INVENTION:  Polypeptides
;  FILE REFERENCE: 784CIP2B
;  CURRENT APPLICATION NUMBER: US/10/037,270
;  CURRENT FILING DATE:  2002-01-04
;  PRIOR APPLICATION NUMBER: 09/552,317
;  PRIOR FILING DATE: 2000-04-25
;  PRIOR APPLICATION NUMBER: 09/488,725
;  PRIOR FILING DATE: 2000-01-21
;  NUMBER OF SEQ ID NOS: 1104
;  SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 773
;   LENGTH: 2629
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (73)..(798)
US-10-037-270-773

  Query Match            20.4%;  Score 425.2;  DB 15;  Length 2629;
  Best Local Similarity   74.1%;  Pred. No. 5.4e-97;
  Matches  557;  Conservative    0;  Mismatches  183;  Indels   12;  Gaps    1;

Qy        175 CGCCCGCCTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGA 234
              |||  | | |||||  ||||||||  |||| ||||||||||||| || || ||| || ||
Db         57 CGCGGGTCAGGACACAATGTTTGCACGAGGACTGAAGAGGAAATGTGTTGGCCACGAGGA 116

Qy        235 AGGAGTAGAGGGT----------TTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACA 282
              ||  || |||||             ||    || ||  | || || |||||||||| ||
Db        117 AGACGTGGAGGGAGCCCTGGCCGGCTTGAAGACAGTGTCCTCATACAGCCTGCAGCGGCA 176

Qy        283 GTCACTCCTGGACATGTCCCTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAA 342
              ||| |||||||||||||| || || ||| | ||||| || |||||||| ||||||||||
Db        177 GTCGCTCCTGGACATGTCTCTGGTGAAGTTGCAGCTTTGCCACATGCTTGTGGAGCCCAA 236

Qy        343 TCTCTGCCGCTCGGTCCTCATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCA 402
              || |||||||| |||||||| |||||||| |||||||||||||| ||||| ||||  ||
Db        237 CCTGTGCCGCTCAGTCCTCATTGCCAACACGGTCCGGCAGATCCAAGAGGAGATGACGCA 296

Qy        403 GGATGGTGTGTGGCATGGGATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCT 462
              ||||||   |||||      |||||||||||   || ||| ||||| ||  | || ||| |
Db        297 GGATGGGACGTGGCGCACAGTGGCACCCCAGGCTGCAGAGCGGGCGCCGCTCGACCGCTT 356

Qy        463 GGTGTCCACAGAGATCCTGTGTCGTACAGTGAGGGGAGCTGAGGAAGAGCACCCTGCTCC 522
              ||| ||||| |||||||||||| ||| ||| | ||||     |||| |  || |||||| |
Db        357 GGTCTCCACGGAGATCCTGTGCCGTGCAGCGTGGGGGCAAGAGGGGGCACATCCTGCTTC 416

Qy        523 TGAACTGGAAGATGCTCCCTTGCAAAACTCGGTTTCCGAGCTCCCCATCGTTGGCTCAGC 582
              ||   |||  || |  | |   ||    | ||||| || ||    |   ||   ||||||
Db        417 TGGCTTGGGGGACGGCCACACACAGGGTCCAGTTTCTGACCTTTGCCCAGTCACCTCAGC 476

Qy        583 ACCAGGGCAAAGGAACCCTCAGAGCAGCCTCTGGGAGATGGACAGCCCACAAGAAAACAG 642
              ||  |  | |||| |||  |||||||||  ||||||||||||  |||| | |||||||||

Db        477 ACAGGCACCAAGGCACCTGCAGAGCAGCGCCTGGGAGATGGATGGCCCTCGAGAAAACAG 536

Qy        643 GGGAAGCTTTCAGAAGTCACTGGACCAGATATTTGAGACCCTGGAGAACAAAAACTCCAG 702
              |||||||||||| |||||||| || |||||||||||| || |||||||  |||||| ||||
Db        537 AGGAAGCTTTCACAAGTCACTTGATCAGATATTTGAAACGCTGGAGACTAAAAACCCCAG 596

Qy        703 TTCAGTGGAGGAACTCTTCTCAGATGTGGACAGCTCCTACTATGACCTGGACACAGTGCT 762
              |    |||| || || |||||||| ||||||||| ||||||| ||||||||||||||| ||
Db        597 CTGCATGGAAGAGCTGTTCTCAGACGTGGACAGCCCCTACTACGACCTGGACACAGTACT 656

Qy        763 AACAGGAATGATGAGTGGGACCAAGTCCAGTCTCTGCAATGGCCTTGAGGGCTTTGCTGC 822
              ||||| |||||| | ||  ||| | |  | | |||| | || || ||||||||| ||| |
Db        657 GACAGGCATGATGGGGGGTGCCAGGCCGGGCCCCTGCGAAGGGCTCGAGGGCTTGGCTCC 716

Qy        823 AGCCACCCCTCCTCCCAGTTCCACTTGCAAGTCTGACCTGGCTGAGCTGGACCATGTGGT 882
              ||||||||    || || |||| |||||||| |||||||  |||||||||| |||||
Db        717 GGCCACCCCAGGCCCTAGCTCCAGCTGCAAGTCCGACCTGGGCGAGCTGGACCACGTGGT 776

Qy        883 AGAGATTCTGGTGGAGACCTGAGAGGCCACCC 914
              ||||| ||||||||||||||||  |   |||
Db        777 GGAGATCCTGGTGGAGACCTGAGCAGGAGCCC 808


RESULT 8
US-10-117-722-773
; Sequence 773, Application US/10117722
; Publication No. US20030219744A1
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi, Vinod
;  APPLICANT:  Zhang, Jie
;  APPLICANT:  Drmanac, Radoje T.
;  TITLE OF INVENTION: No. US20030219744A1el Nucleic Acids and
;  TITLE OF INVENTION:  Polypeptides
;  FILE REFERENCE: 784CIP2BCIP
;  CURRENT APPLICATION NUMBER: US/10/117,722
;  CURRENT FILING DATE:  2002-04-04
;  PRIOR APPLICATION NUMBER: 09/620,312
;  PRIOR FILING DATE: 2000-07-19
;  PRIOR APPLICATION NUMBER: 09/552,317
;  PRIOR FILING DATE: 2000-04-25
;  PRIOR APPLICATION NUMBER: 09/488,725
;  PRIOR FILING DATE: 2000-01-21
;  NUMBER OF SEQ ID NOS: 1104
;  SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 773
;   LENGTH: 2629
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (73)..(798)
US-10-117-722-773

  Query Match            20.4%;  Score 425.2;  DB 16;  Length 2629;
  Best Local Similarity   74.1%;  Pred. No. 5.4e-97;
  Matches  557;  Conservative    0;  Mismatches  183;  Indels   12;  Gaps    1;

Qy        175 CGCCCGCCTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGA 234
              |||  | | |||||  ||||||||  |||| ||||||||||||| || || ||| || ||
Db         57 CGCGGGTCAGGACACAATGTTTGCACGAGGACTGAAGAGGAAATGTGTTGGCCACGAGGA 116

Qy        235 AGGAGTAGAGGGT----------TTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACA 282
              ||  || |||||             ||    || ||  | || || |||||||||| ||
Db        117 AGACGTGGAGGGAGCCCTGGCCGGCTTGAAGACAGTGTCCTCATACAGCCTGCAGCGGCA 176

Qy        283 GTCACTCCTGGACATGTCCCTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAA 342
              ||| |||||||||||||| || || ||| | ||||| || |||||||| ||||||||||
Db        177 GTCGCTCCTGGACATGTCTCTGGTGAAGTTGCAGCTTTGCCACATGCTTGTGGAGCCCAA 236

Qy        343 TCTCTGCCGCTCGGTCCTCATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCA 402
              || |||||||| |||||||| |||||||| |||||||||||||| ||||| ||||  ||
```

```
Db        237 CCTGTGCCGCTCAGTCCTCATTGCCAACACGGTCCGGCAGATCCAAGAGGAGATGACGCA 296

Qy        403 GGATGGTGTGTGGCATGGGATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCT 462

Db        297 GGATGGGACGTGGCGCACAGTGGCACCCCAGGCTGCAGAGCGGGCGCCGCTCGACCGCTT 356

Qy        463 GGTGTCCACAGAGATCCTGTGTCGTACAGTGAGGGGAGCTGAGGAAGAGCACCCTGCTCC 522

Db        357 GGTCTCCACGGAGATCCTGTGCCGTGCAGCGTGGGGGCAAGAGGGGGCACATCCTGCTTC 416

Qy        523 TGAACTGGAAGATGCTCCCTTGCAAAACTCGGTTTCCGAGCTCCCCATCGTTGGCTCAGC 582

Db        417 TGGCTTGGGGGACGGCCACACACAGGGTCCAGTTTCTGACCTTTGCCCAGTCACCTCAGC 476

Qy        583 ACCAGGGCAAAGGAACCCTCAGAGCAGCCTCTGGGAGATGGACAGCCCACAAGAAAACAG 642

Db        477 ACAGGCACCAAGGCACCTGCAGAGCAGCGCCTGGGAGATGGATGGCCCTCGAGAAAACAG 536

Qy        643 GGGAAGCTTTCAGAAGTCACTGGACCAGATATTTGAGACCCTGGAGAACAAAAACTCCAG 702

Db        537 AGGAAGCTTTCACAAGTCACTTGATCAGATATTTGAAACGCTGGAGACTAAAAACCCCAG 596

Qy        703 TTCAGTGGAGGAACTCTTCTCAGATGTGGACAGCTCCTACTATGACCTGGACACAGTGCT 762

Db        597 CTGCATGGAAGAGCTGTTCTCAGACGTGGACAGCCCCTACTACGACCTGGACACAGTACT 656

Qy        763 AACAGGAATGATGAGTGGGACCAAGTCCAGTCTCTGCAATGGCCTTGAGGGCTTTGCTGC 822

Db        657 GACAGGCATGATGGGGGGTGCCAGGCCGGGCCCCTGCGAAGGGCTCGAGGGCTTGGCTCC 716

Qy        823 AGCCACCCCTCCTCCCAGTTCCACTTGCAAGTCTGACCTGGCTGAGCTGGACCATGTGGT 882

Db        717 GGCCACCCCAGGCCCTAGCTCCAGCTGCAAGTCCGACCTGGGCGAGCTGGACCACGTGGT 776

Qy        883 AGAGATTCTGGTGGAGACCTGAGAGGCCACCC 914

Db        777 GGAGATCCTGGTGGAGACCTGAGCAGGAGCCC 808


RESULT 9
US-09-825-294-167
; Sequence 167, Application US/09825294
; Patent No. US20020004491A1
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Stolk, John A.
;  APPLICANT:  Algate, Paul A.
;  APPLICANT:  Fling, Steven P.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
;  TITLE OF INVENTION:  THERAPY AND DIAGNOSIS OF OVARIAN CANCER
;  FILE REFERENCE: 210121.484C5
;  CURRENT APPLICATION NUMBER: US/09/825,294
;  CURRENT FILING DATE:  2001-04-03
;  NUMBER OF SEQ ID NOS: 215
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 167
;   LENGTH: 396
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-825-294-167

  Query Match             9.1%;  Score 190;  DB 9;  Length 396;
  Best Local Similarity   74.7%;  Pred. No. 1.1e-37;
  Matches  257;  Conservative    0;  Mismatches   75;  Indels   12;  Gaps    1;

Qy        175 CGCCCGCCTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGA 234

Db         52 CGCGGGTCAGGACACAATGTTTGCACGAGGACTGAAGAGGAAATGTGTTGGCCACGAGGA 111

Qy        235 AGGAGTAGAGGGT------------TTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACA 282

Db        112 AGACGTGGAGGGAGCCCTGGCCGGCTTGAAGACAGTGTCCTCATACAGCCTGCAGCGGCA 171

Qy        283 GTCACTCCTGGACATGTCCCTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAA 342

Db        172 GTCGCTCCTGGACATGTCTCTGGTGAAGTTGCAGCTTTGCCACATGCTTGTGGAGCCCAA 231

Qy        343 TCTCTGCCGCTCGGTCCTCATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCA 402

Db        232 TCTGTGCCGCTCAGTCCTCATTGCCAACACGGTCCGGCAGATCCAAGAGGAGATGACGCA 291

Qy        403 GGATGGTGTGTGGCATGGGATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCT 462

Db        292 GGATGGGACGTGGCGCACAGTGGCACCCCAGGCTGCAGAGCGGGCGCCGCTCGACCGCTT 351

Qy        463 GGTGTCCACAGAGATCCTGTGTCGTACAGTGAGGGGAGCTGAGG 506

Db        352 GGTCTCCACGGAGATCCTGTGCCGTGCAGCGTGGGGGCAAGAGG 395


RESULT 10
US-09-970-966-167
; Sequence 167, Application US/09970966
; Patent No. US20020173638A1
; GENERAL INFORMATION:
;  APPLICANT: Stolk, John A.
;  APPLICANT:  Molesh, David Alan
;  APPLICANT:  Fling, Steven P.
;  APPLICANT:  Xu, Jiangchun
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF OVARIAN CANCER
;  FILE REFERENCE: 210121.484C6
;  CURRENT APPLICATION NUMBER: US/09/970,966
;  CURRENT FILING DATE:  2001-10-02
;  NUMBER OF SEQ ID NOS: 215
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 167
;   LENGTH: 396
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-970-966-167

  Query Match             9.1%;  Score 190;  DB 9;  Length 396;
  Best Local Similarity   74.7%;  Pred. No. 1.1e-37;
  Matches  257;  Conservative    0;  Mismatches   75;  Indels   12;  Gaps    1;

Qy        175 CGCCCGCCTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGA 234

Db         52 CGCGGGTCAGGACACAATGTTTGCACGAGGACTGAAGAGGAAATGTGTTGGCCACGAGGA 111

Qy        235 AGGAGTAGAGGGT------------TTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACA 282

Db        112 AGACGTGGAGGGAGCCCTGGCCGGCTTGAAGACAGTGTCCTCATACAGCCTGCAGCGGCA 171

Qy        283 GTCACTCCTGGACATGTCCCTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAA 342

Db        172 GTCGCTCCTGGACATGTCTCTGGTGAAGTTGCAGCTTTGCCACATGCTTGTGGAGCCCAA 231

Qy        343 TCTCTGCCGCTCGGTCCTCATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCA 402

Db        232 TCTGTGCCGCTCAGTCCTCATTGCCAACACGGTCCGGCAGATCCAAGAGGAGATGACGCA 291

Qy        403 GGATGGTGTGTGGCATGGGATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCT 462

Db        292 GGATGGGACGTGGCGCACAGTGGCACCCCAGGCTGCAGAGCGGGCGCCGCTCGACCGCTT 351

Qy        463 GGTGTCCACAGAGATCCTGTGTCGTACAGTGAGGGGAGCTGAGG 506

Db        352 GGTCTCCACGGAGATCCTGTGCCGTGCAGCGTGGGGGCAAGAGG 395


RESULT 11
US-10-212-677-167
; Sequence 167, Application US/10212677
; Publication No. US20030129192A1
; GENERAL INFORMATION:
```

```
;  APPLICANT: Chenault, Ruth A.
;  APPLICANT:  Xu, Jiangchun
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Harlocker, Susan L.
;  APPLICANT:  McNeill, Patricia D.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF OVARIAN CANCER
;  FILE REFERENCE: 210121.484C7
;  CURRENT APPLICATION NUMBER: US/10/212,677
;  CURRENT FILING DATE:  2002-08-02
;  NUMBER OF SEQ ID NOS: 288
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 167
;   LENGTH: 396
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-212-677-167

  Query Match             9.1%;  Score 190;  DB 15;  Length 396;
  Best Local Similarity   74.7%;  Pred. No. 1.1e-37;
  Matches  257;  Conservative    0;  Mismatches   75;  Indels   12;  Gaps    1;

Qy        175 CGCCCGCCTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGA 234
              |||  | | |||||  ||||||||  |||| ||||||||||||| || || ||| || ||
Db         52 CGCGGGTCAGGACACAATGTTTGCACGAGGACTGAAGAGGAAATGTGTTGGCCACGAGGA 111

Qy        235 AGGAGTAGAGGGT------------TTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACA 282
              ||  || |||||              ||    || ||  | || || ||||||||||| ||
Db        112 AGACGTGGAGGGAGCCCTGGCCGGCTTGAAGACAGTGTCCTCATACAGCCTGCAGCGGCA 171

Qy        283 GTCACTCCTGGACATGTCCCTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAA 342
              ||| ||||||||||||||| || || ||| | ||||| || |||||||| ||||||||||
Db        172 GTCGCTCCTGGACATGTCTCTGGTGAAGTTGCAGCTTTGCCACATGCTTGTGGAGCCCAA 231

Qy        343 TCTCTGCCGCTCGGTCCTCATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCA 402
              ||| |||||||| |||||||| |||||||| |||||||||||||| ||||| ||||  ||
Db        232 TCTGTGCCGCTCAGTCCTCATTGCCAACACGGTCCGGCAGATCCAAGAGGAGATGACGCA 291

Qy        403 GGATGGTGTGTGGCATGGGATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCT 462
              ||||||   |||||      ||||||||||  || ||| ||||| ||  | || ||| |
Db        292 GGATGGGACGTGGCGCACAGTGGCACCCCAGGCTGCAGAGCGGGCGCCGCTCGACCGCTT 351

Qy        463 GGTGTCCACAGAGATCCTGTGTCGTACAGTGAGGGGAGCTGAGG 506
              ||| ||||| |||||||||| ||| ||| | ||||    ||||
Db        352 GGTCTCCACGGAGATCCTGTGCCGTGCAGCGTGGGGGCAAGAGG 395


RESULT 12
US-10-361-811-167
; Sequence 167, Application US/10361811
; Publication No. US20030206918A1
; GENERAL INFORMATION:
;  APPLICANT: Fanger, Gary R.
;  APPLICANT:  Fling, Steven P.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF OVARIAN CANCER
;  FILE REFERENCE: 210121.484C8
;  CURRENT APPLICATION NUMBER: US/10/361,811
;  CURRENT FILING DATE:  2003-02-05
;  NUMBER OF SEQ ID NOS: 293
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 167
;   LENGTH: 396
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-361-811-167

  Query Match             9.1%;  Score 190;  DB 16;  Length 396;
  Best Local Similarity   74.7%;  Pred. No. 1.1e-37;
  Matches  257;  Conservative    0;  Mismatches   75;  Indels   12;  Gaps    1;

Qy        175 CGCCCGCCTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGA 234
              |||  | | |||||  ||||||||  |||| ||||||||||||| || || ||| || ||
Db         52 CGCGGGTCAGGACACAATGTTTGCACGAGGACTGAAGAGGAAATGTGTTGGCCACGAGGA 111

Qy        235 AGGAGTAGAGGGT------------TTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACA 282
              ||  || |||||              ||    || ||  | || || ||||||||||| ||
Db        112 AGACGTGGAGGGAGCCCTGGCCGGCTTGAAGACAGTGTCCTCATACAGCCTGCAGCGGCA 171

Qy        283 GTCACTCCTGGACATGTCCCTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAA 342
              ||| ||||||||||||||| || || ||| | ||||| || |||||||| ||||||||||
Db        172 GTCGCTCCTGGACATGTCTCTGGTGAAGTTGCAGCTTTGCCACATGCTTGTGGAGCCCAA 231

Qy        343 TCTCTGCCGCTCGGTCCTCATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCA 402
              ||| |||||||| |||||||| |||||||| |||||||||||||| ||||| ||||  ||
Db        232 TCTGTGCCGCTCAGTCCTCATTGCCAACACGGTCCGGCAGATCCAAGAGGAGATGACGCA 291

Qy        403 GGATGGTGTGTGGCATGGGATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCT 462
              ||||||   |||||      ||||||||||  || ||| ||||| ||  | || ||| |
Db        292 GGATGGGACGTGGCGCACAGTGGCACCCCAGGCTGCAGAGCGGGCGCCGCTCGACCGCTT 351

Qy        463 GGTGTCCACAGAGATCCTGTGTCGTACAGTGAGGGGAGCTGAGG 506
              ||| ||||| |||||||||| ||| ||| | ||||    ||||
Db        352 GGTCTCCACGGAGATCCTGTGCCGTGCAGCGTGGGGGCAAGAGG 395


RESULT 13
US-10-369-186-167
; Sequence 167, Application US/10369186
; Publication No. US20030232056A1
; GENERAL INFORMATION:
;  APPLICANT: Fanger, Gary R.
;  APPLICANT:  Fling, Steven P.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF OVARIAN CANCER
;  FILE REFERENCE: 210121.484C9
;  CURRENT APPLICATION NUMBER: US/10/369,186
;  CURRENT FILING DATE:  2003-02-14
;  NUMBER OF SEQ ID NOS: 293
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 167
;   LENGTH: 396
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-369-186-167

  Query Match             9.1%;  Score 190;  DB 16;  Length 396;
  Best Local Similarity   74.7%;  Pred. No. 1.1e-37;
  Matches  257;  Conservative    0;  Mismatches   75;  Indels   12;  Gaps    1;

Qy        175 CGCCCGCCTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGA 234
              |||  | | |||||  ||||||||  |||| ||||||||||||| || || ||| || ||
Db         52 CGCGGGTCAGGACACAATGTTTGCACGAGGACTGAAGAGGAAATGTGTTGGCCACGAGGA 111

Qy        235 AGGAGTAGAGGGT------------TTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACA 282
              ||  || |||||              ||    || ||  | || || ||||||||||| ||
Db        112 AGACGTGGAGGGAGCCCTGGCCGGCTTGAAGACAGTGTCCTCATACAGCCTGCAGCGGCA 171

Qy        283 GTCACTCCTGGACATGTCCCTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAA 342
              ||| ||||||||||||||| || || ||| | ||||| || |||||||| ||||||||||
Db        172 GTCGCTCCTGGACATGTCTCTGGTGAAGTTGCAGCTTTGCCACATGCTTGTGGAGCCCAA 231

Qy        343 TCTCTGCCGCTCGGTCCTCATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCA 402
              ||| |||||||| |||||||| |||||||| |||||||||||||| ||||| ||||  ||
Db        232 TCTGTGCCGCTCAGTCCTCATTGCCAACACGGTCCGGCAGATCCAAGAGGAGATGACGCA 291

Qy        403 GGATGGTGTGTGGCATGGGATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCT 462
              ||||||   |||||      ||||||||||  || ||| ||||| ||  | || ||| |
Db        292 GGATGGGACGTGGCGCACAGTGGCACCCCAGGCTGCAGAGCGGGCGCCGCTCGACCGCTT 351

Qy        463 GGTGTCCACAGAGATCCTGTGTCGTACAGTGAGGGGAGCTGAGG 506
              ||| ||||| |||||||||| ||| ||| | ||||    ||||
Db        352 GGTCTCCACGGAGATCCTGTGCCGTGCAGCGTGGGGGCAAGAGG 395


RESULT 14
```

```
US-09-917-800A-1285/c
; Sequence 1285, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
;  APPLICANT: Mendrick, Donna
;  APPLICANT:  Porter, Mark
;  APPLICANT:  Johnson, Kory
;  APPLICANT:  Castle, Arthur
;  APPLICANT:  Elashoff, Michael
;  APPLICANT:  Gene Logic, Inc.
;  TITLE OF INVENTION: Molecular Toxicology Modeling
;  FILE REFERENCE: 44921-5038-US
;  CURRENT APPLICATION NUMBER: US/09/917,800A
;  CURRENT FILING DATE:  2001-07-31
;  PRIOR APPLICATION NUMBER: US 60/222,040
;  PRIOR FILING DATE: 2000-07-31
;  PRIOR APPLICATION NUMBER: US 60/222,880
;  PRIOR FILING DATE: 2000-11-02
;  PRIOR APPLICATION NUMBER: US 60/290,029
;  PRIOR FILING DATE: 2001-05-11
;  PRIOR APPLICATION NUMBER: US 60/290,645
;  PRIOR FILING DATE: 2001-05-15
;  PRIOR APPLICATION NUMBER: US 60/292,336
;  PRIOR FILING DATE: 2001-05-22
;  PRIOR APPLICATION NUMBER: US 60/295,798
;  PRIOR FILING DATE: 2001-06-06
;  PRIOR APPLICATION NUMBER: US 60/297,457
;  PRIOR FILING DATE: 2001-06-13
;  PRIOR APPLICATION NUMBER: US 60/298,884
;  PRIOR FILING DATE: 2001-06-19
;  PRIOR APPLICATION NUMBER: US 60/303,459
;  PRIOR FILING DATE: 2001-07-09
;  NUMBER OF SEQ ID NOS: 1740
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1285
;   LENGTH: 522
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. US20020119462A1 AI236771
US-09-917-800A-1285

  Query Match             8.9%;  Score 185.4;  DB 9;  Length 522;
  Best Local Similarity   73.8%;  Pred. No. 2e-36;
  Matches  401;  Conservative    0;  Mismatches   91;  Indels   51;  Gaps   11;

Qy       1532 GCTAAGGTCTTACTCCTGGAAATCCCAACACCAGTTCTTCAGGGACTGCTGTGAGGCAGT 1591
             | |||||||||    |||||||| || ||||| |||||||| |||||  || |||||||
Db        519 GTTAAGGTCTTCAGTCTGGAAATCACAGCACCA-TTCTTCAGAGACTGTGGTCAGGCAGT 461

Qy       1592 GCCTTATGCAGGTCTTGTCCTTGGCCATCACTGTCTGGTTCCCAGCCCAGCAC-ATGTGA 1650
             ||||||||||||||| |||||||             ||| ||||||||||||  |||||
Db        460 GCCTTATGCAGGTCTGGTCCTTG------------GGTCCCCAGCCCAGCACTGTGTGA 414

Qy       1651 CATGAGGACATGACATGCCCGAACCACCCAGCACCACATGCTCCATGTCAAGTGTGTACG 1710
             || ||||||| ||||| ||    | || |||||||||||||| |||||||| || | ||
Db        413 CACGAGGACACGACATCCCAACAACAACCAGCACCACATGCTACATGTCAAGAGTATGCG 354

Qy       1711 TGGAGACCACTGGCTCCCAGGC----CTGTGCTCAGAGAGGGTGTGCAGTCCTACGTGTG 1766
             |||||||  | || | || |||     |||||| ||| |||||||||| ||  | ||||
Db        353 TGGAGACTGCAGGGTGCCGGGCATGTGTGTGCTTAGAAAGGGTGTGCAGCCCACCATGTG 294

Qy       1767 CTGGGGGGGACGACGGTGACCTGTGCTTGCTTGCTTTTAAAATGGTGCTTGGACGTTTTA 1826
             ||       |  | |   |||||||||||||||| |||||||||||||||| || |||||
Db        293 CT------CAATAGGAGTACCTGTGCTTGCTTGC-TTTAAAATGGTGCTTGAACATTTTA 241

Qy       1827 AGGTTAAAAACAATCCGACTCCATATGATTTAGGGCTCCTCCACCCTGGGGTGGCCCCTA 1886
             |||||||||||||| || ||  ||  | |||| |||||||| |||||||||||||| |
Db        240 AGGTTAAAAACAATCTGATTCTGTA--ACTTAGTGCTCCTCCATCCTGGGGTGGCCCC-A 184

Qy       1887 TGCTGTCTGCTTGGATCTCAAAGTCTTGGT---------ACTCGGCACTGTCAGACTCCA 1937
             |||||||||||||| |||||  |||  | |          ||  | ||    |   | |||
Db        183 TGCTGTCTGCTTGGGTCTCATGGTCCCGATAACCAGATCACATGTCAGACCCCACCCCCA 124

Qy       1938 CCCCATGTATCCTTTTTGTTTCTCTTGTGCTTTTTTTGGACTTCCCAACCTGAGCCTA-- 1995
             |||||      | ||| | | |||  ||   || |||||||| |||| ||||||||||
Db        123 CCCCACCATGCTTTTGTTTCTCTTGTGCTGTTCTTTTGGACGTCCCCACCTGAGCCTAAG 64

Qy       1996 AGGTTTTATTTTATATGTG-----------CTTCAATATCAACAATGTAAACCTCACTTT 2044
             |||||||||||| ||||||           ||||| ||||||||||||||||||||||||
Db         63 AGGTTTTATTTTGTATGTGTTATGACTGTTCTTCACTATCAACAATGTAAACCTCACTTT 4

Qy       2045 ATT 2047
             |||
Db           3 ATT 1


RESULT 15
US-09-918-995-12211
; Sequence 12211, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 12211
;   LENGTH: 477
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(477)
;   OTHER INFORMATION: n = A,T,C or G
US-09-918-995-12211

  Query Match             4.7%;  Score 98.2;  DB 10;  Length 477;
  Best Local Similarity   78.9%;  Pred. No. 2.9e-14;
  Matches  157;  Conservative    0;  Mismatches   33;  Indels    9;  Gaps    3;

Qy       1103 ACTTTCTAAAATTAAGTTTTATATGTTTTTGGGCAATATTTTGTCTTAAGATATATTTTT 1162
             | | | || |||| ||||||| ||| ||| ||||| ||||||||||||||||||||||||
Db         61 AATCTTTACAATTCAGTTTTACATG-TTTCGGGCAGTATTTTGTCTTAAGATATATTTTT 119

Qy       1163 TAAACTTTTTATACTTTA-------GATTTTTTTCAGCTATTTTCTTAAAAGTATATTTT 1215
             ||||||||||||||| |||          ||||||||||||||||||||||||||||||
Db        120 TAAACTTTTTATACCTTATCGGTTGAGATTTTTTCAGCTATTTTCTTAAAAGTATATTTT 179

Qy       1216 TTCTACAAACATCCTCTGCTGCTACATTAGAAACATTTATAACCTAAATACGATTGGTGT 1275
             ||||| |||||||||| |||||||||||||| ||| |||||| ||||||      ||
Db        180 TTCTATAAACATCCTTTGCTGCTACATTAG-AACTTTTATAGCCTAAACAATTGCAGTTG 238

Qy       1276 GTCATTTTAAAGGTTTAAA 1294
             ||   ||| |   ||| ||
Db        239 GTGTGTTTCATTTTTTTAA 257


Search completed: September 10, 2004, 17:49:05
Job time : 706 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: September 10, 2004, 10:57:36 ; Search time 5666 Seconds
(without alignments)
10973.007 Million cell updates/sec

Title: US-10-076-069-1
Perfect score: 2082
Sequence: 1 ccccgcgtcggtcttccacc.....atggaaaaaaaaaaaaaaaa 2082

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1883.8	90.5	1985	11	AK010535	AK010535 Mus muscu
	2	1883.8	90.5	1996	11	AK032980	AK032980 Mus muscu
	3	890.2	42.8	1180	13	BU511608	BU511608 AGENCOURT
	4	849.2	40.8	989	9	AV000968	AV000968 AV000968
	5	810.4	38.9	836	13	BU054465	BU054465 UI-M-FD0-
	6	809.8	38.9	919	14	CB183964	CB183964 AGENCOURT
	7	798	38.3	927	13	BQ950078	BQ950078 AGENCOURT
	8	790.6	38.0	920	13	BU148336	BU148336 AGENCOURT
	9	785.8	37.7	926	13	BU151870	BU151870 AGENCOURT
	10	755.4	36.3	994	10	BB609465	BB609465 BB609465
	11	748.4	35.9	967	13	BQ936878	BQ936878 AGENCOURT
	12	746	35.8	922	13	BQ713776	BQ713776 AGENCOURT
	13	741	35.6	830	12	BI661691	BI661691 603305632
	14	721.4	34.6	800	12	BI153019	BI153019 602918211
	15	714	34.3	714	29	AY403475	AY403475 Mus muscu
	16	697.8	33.5	809	13	BQ572821	BQ572821 UI-M-FD0-
	17	694.8	33.4	826	13	BU706179	BU706179 UI-M-FR0-
	18	687.8	33.0	815	14	CA752039	CA752039 UI-M-FD0-
	19	681.8	32.7	824	14	CA752149	CA752149 UI-M-FD0-
	20	677.8	32.6	748	12	BI689943	BI689943 603310777
	21	677.8	32.6	876	12	BI685217	BI685217 603310227
	22	677	32.5	753	12	BI685932	BI685932 603309232
	23	676.4	32.5	936	13	BU512943	BU512943 AGENCOURT
	24	672.8	32.3	911	13	BQ886760	BQ886760 AGENCOURT
	25	669.4	32.2	767	12	BI653556	BI653556 603300591
c	26	668	32.1	679	10	BG069072	BG069072 H3072D09-
	27	665.6	32.0	713	12	BI665883	BI665883 603286912
	28	664.2	31.9	682	10	BB055758	BB055758 BB055758
	29	661.6	31.8	914	14	CB587874	CB587874 AGENCOURT
	30	645.8	31.0	649	14	CB057494	CB057494 NISC_js05
	31	636.6	30.6	724	14	CA751730	CA751730 UI-M-FD0-
	32	620.4	29.8	898	13	BU515897	BU515897 AGENCOURT
	33	594.6	28.6	724	10	BF607870	BF607870 MY1_00081
	34	588.8	28.3	679	13	BY726193	BY726193 BY726193
	35	585.6	28.1	804	10	BE573436	BE573436 601332611
	36	573.8	27.6	648	14	CF173487	CF173487 B0924D06-
c	37	573	27.5	730	12	BI294797	BI294797 UI-R-DK0-
	38	571.8	27.5	646	14	CF916191	CF916191 B0991D06-
	39	568.2	27.3	726	12	BI077734	BI077734 602871683
c	40	564.4	27.1	589	10	AW547757	AW547757 L0026D04-
	41	552	26.5	552	12	BI663600	BI663600 603288096
	42	548.2	26.3	906	10	BF144504	BF144504 601790146
	43	544.4	26.1	600	14	CA528387	CA528387 8072-96 M
	44	528.6	25.4	568	12	BI666485	BI666485 603288772
	45	525	25.2	613	29	CG586964	CG586964 OST236075

ALIGNMENTS

RESULT 1
AK010535
LOCUS AK010535 1985 bp mRNA linear HTC 20-SEP-2003
DEFINITION Mus musculus ES cells cDNA, RIKEN full-length enriched library, clone:2410018C03 product:HEMATOPOIETIC PROGENITOR PROTEIN, full insert sequence.
ACCESSION AK010535
VERSION AK010535.1 GI:12846052
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159

REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 1985)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI. Host: SOLR.

FEATURES Location/Qualifiers
source 1. .1985
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:2410018C03"
/db_xref="MGI:1904673"
/db_xref="taxon:10090"
/clone="2410018C03"
/cell_type="ES cells"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
CDS 115. .828
/note="unnamed protein product; HEMATOPOIETIC PROGENITOR PROTEIN (SPTR|Q99MP6, evidence: FASTY, 100%ID, 100%length, match=711)
putative"
/codon_start=1
/protein_id="BAB27012.2"
/db_xref="GI:26369635"
/translation="MFARGLKRKYGDQEEGVEGFGTVPSYSLQRQSLLDMSLVKLQLC
HMLVEPNLCRSVLIANTVRQIQEEMSQDGVWHGMAPQNVDRAPVERLVSTEILCRTVR
GAEEEHPAPELEDAPLQNSVSELPIVGSAPGQRNPQSSLWEMDSPQENRGSFQKSLDQ
IFETLENKNSSSVEELFSDVDSSYYDLDTVLTGMMSGTKSSLCNGLEGFAAATPPPSS
TCKSDLAELDHVVEILVET"
polyA_signal 1969. .1974
/note="putative"
polyA_site 1985
/note="putative"

ORIGIN

Query Match 90.5%; Score 1883.8; DB 11; Length 1985;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1878; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      182 CTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGAAGGAGTA 241
Db      106 CCGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGAAGGAGTA 165

Qy      242 GAGGGTTTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACAGTCACTCCTGGACATGTCC 301
Db      166 GAGGGTTTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACAGTCACTCCTGGACATGTCC 225

Qy      302 CTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAATCTCTGCCGCTCGGTCCTC 361
Db      226 CTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAATCTCTGCCGCTCGGTCCTC 285

Qy      362 ATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCAGGATGGTGTGTGGCATGGG 421
Db      286 ATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCAGGATGGTGTGTGGCATGGG 345

Qy      422 ATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCTGGTGTCCACAGAGATCCTG 481
Db      346 ATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCTGGTGTCCACAGAGATCCTG 405

Qy      482 TGTCGTACAGTGAGGGGAGCTGAGGAAGAGCACCCTGCTCCTGAACTGGAAGATGCTCCC 541
Db      406 TGTCGTACAGTGAGGGGAGCTGAGGAAGAGCACCCTGCTCCTGAACTGGAAGATGCTCCC 465

Qy      542 TTGCAAAACTCGGTTTCCGAGCTCCCCATCGTTGGCTCAGCACCAGGGCAAAGGAACCCT 601
Db      466 TTGCAAAACTCGGTTTCCGAGCTCCCCATCGTTGGCTCAGCACCAGGGCAAAGGAACCCT 525

Qy      602 CAGAGCAGCCTCTGGGAGATGGACAGCCCACAAGAAAACAGGGGAAGCTTTCAGAAGTCA 661
Db      526 CAGAGCAGCCTCTGGGAGATGGACAGCCCACAAGAAAACAGGGGAAGCTTTCAGAAGTCA 585

Qy      662 CTGGACCAGATATTTGAGACCCTGGAGAACAAAAACTCCAGTTCAGTGGAGGAACTCTTC 721
Db      586 CTGGACCAGATATTTGAGACCCTGGAGAACAAAAACTCCAGTTCAGTGGAGGAACTCTTC 645

Qy      722 TCAGATGTGGACAGCTCCTACTATGACCTGGACACAGTGCTAACAGGAATGATGAGTGGG 781
Db      646 TCAGATGTGGACAGCTCCTACTATGACCTGGACACAGTGCTAACAGGAATGATGAGTGGG 705

Qy      782 ACCAAGTCCAGTCTCTGCAATGGCCTTGAGGGCTTTGCTGCAGCCACCCCTCCTCCCAGT 841
Db      706 ACCAAGTCCAGTCTCTGCAATGGCCTTGAGGGCTTTGCTGCAGCCACCCCTCCTCCCAGT 765

Qy      842 TCCACTTGCAAGTCTGACCTGGCTGAGCTGGACCATGTGGTAGAGATTCTGGTGGAGACC 901
Db      766 TCCACTTGCAAGTCTGACCTGGCTGAGCTGGACCATGTGGTAGAGATTCTGGTGGAGACC 825

Qy      902 TGAGAGGCCACCCCAGTGGGCTAAGGGTGAGGCCACCAGTCCCCATGGAGCTCACGTGTG 961
Db      826 TGAGAGGCCACCCCAGTGGGCTAAGGGTGAGGCCACCAGTCCCCATGGAGCTCACGTGTG 885

Qy      962 TTGTGACCCAGAGACAGATAAGCACTTGTCCTAAGAGGGGCTCTGGCTCTTGAGCTCATT 1021
```

```
Db        886 TTGTGACCCAGAGACAGATAAGCACTTGTCCTAAGAGGGGCTCTGGCTCTTGAGCTCATT 945

Qy       1022 ATCCTTTTGTGTGACATTGGACTCACTGTGGAGGATGGTGTGTCACAGCTATGTCTAGTC 1081
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        946 ATCCTTTTGTGTGACATTGGACTCACTGTGGAGGATGGTGTGTCACAGCTATGTCTAGTC 1005

Qy       1082 TATTTTCAATTAGATAGGTGAACTTTCTAAAATTAAGTTTTATATGTTTTTGGGCAATAT 1141
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1006 TATTTTCAATTAGATAGGTGAACTTTCTAAAATTAAGTTTTATATGTTTTTGGGCAATAT 1065

Qy       1142 TTTGTCTTAAGATATATTTTTTAAACTTTTTATACTTTAGATTTTTTTCAGCTATTTTCT 1201
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1066 TTTGTCTTAAGATATATTTTTTAAACTTTTTATACTTTAGATTTTTTTCAGCTATTTTCT 1125

Qy       1202 TAAAAGTATATTTTTTCTACAAACATCCTCTGCTGCTACATTAGAAACATTTATAACCTA 1261
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1126 TAAAAGTATATTTTTTCTACAAACATCCTCTGCTGCTACATTAGAAACATTTATAACCTA 1185

Qy       1262 AATACGATTGGTGTGTCATTTTAAAGGTTTAAATAGAAAACTTCTTTTGTTACTGAGTCT 1321
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1186 AATACGATTGGTGTGTCATTTTAAAGGTTTAAATAGAAAACTTCTTTTGTTACTGAGTCT 1245

Qy       1322 CTACACTCCCAAGGCAACTGTAAATGTAGCCGGCCGGGTGTTTACATGAGAGGCTCCAGT 1381
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1246 CTACACTCCCAAGGCAACTGTAAATGTAGCCGGCCGGGTGTTTACATGAGAGGCTCCAGT 1305

Qy       1382 ATGGTCTACATTCTAGTAGAGCTTGAAAAGAACCATGCACAGCTCCACTGCCCCCTCACT 1441
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1306 ATGGTCTACATTCTAGTAGAGCTTGAAAAGAACCATGCACAGCTCCACTGCCCCCTCACT 1365

Qy       1442 GGGTCTGCTCTGGCGGATCGGAGCTCTCTTCCTAGCCCCGTGTGCAGGATGGCTTTATTT 1501
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1366 GGGTCTGCTCTGGCGGATCGGAGCTCTCTTCCTAGCCCCGTGTGCAGGATGGCTTTATTT 1425

Qy       1502 ATGCCTATTTATATGTAAATGCCACTGAAAGCTAAGGTCTTACTCCTGGAAATCCCAACA 1561
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1426 ATGCCTATTTATATGTAAATGCCACTGAAAGCTAAGGTCTTACTCCTGGAAATCCCAACA 1485

Qy       1562 CCAGTTCTTCAGGGACTGCTGTGAGGCAGTGCCTTATGCAGGTCTTGTCCTTGGCCATCA 1621
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1486 CCAGTTCTTCAGGGACTGCTGTGAGGCAGTGCCTTATGCAGGTCTTGTCCTTGGCCATCA 1545

Qy       1622 CTGTCTGGTTCCCAGCCCAGCACATGTGACATGAGGACATGACATGCCCGAACCACCCAG 1681
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1546 CTGTCTGGTTCCCAGCCCAGCACATGTGACATGAGGACATGACATGCCCGAACCACCCAG 1605

Qy       1682 CACCACATGCTCCATGTCAAGTGTGTACGTGGAGACCACTGGCTCCCAGGCCTGTGCTCA 1741
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1606 CACCACATGCTCCATGTCAAGTGTGTACGTGGAGACCACTGGCTCCCAGGCCTGTGCTCA 1665

Qy       1742 GAGAGGGTGTGCAGTCCTACGTGTGCTGGGGGGGACGACGGTGACCTGTGCTTGCTTGCT 1801
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1666 GAGAGGGTGTGCAGTCCTACGTGTGCTGGGGGGGACGACGGTGACCTGTGCTTGCTTGCT 1725

Qy       1802 TTTAAAATGGTGCTTGGACGTTTTAAGGTTAAAAACAATCCGACTCCATATGATTTAGGG 1861
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1726 TTTAAAATGGTGCTTGGACGTTTTAAGGTTAAAAACAATCCGACTCCATATGATTTAGGG 1785

Qy       1862 CTCCTCCACCCTGGGGTGGCCCCTATGCTGTCTGCTTGGATCTCAAAGTCTTGGTACTCG 1921
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1786 CTCCTCCACCCTGGGGTGGCCCCTATGCTGTCTGCTTGGATCTCAAAGTCTTGGTACTCG 1845

Qy       1922 GCACTGTCAGACTCCACCCCATGTATCCTTTTTGTTTCTCTTGTGCTTTTTTTGGACTTC 1981
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1846 GCACTGTCAGACTCCACCCCATGTATCCTTTTTGTTTCTCTTGTGCTTTTTTTGGACTTC 1905

Qy       1982 CCAACCTGAGCCTAAGGTTTTATTTTATATGTGCTTCAATATCAACAATGTAAACCTCAC 2041
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1906 CCAACCTGAGCCTAAGGTTTTATTTTATATGTGCTTCAATATCAACAATGTAAACCTCAC 1965

Qy       2042 TTTATTAAAAGTATCCAGC 2060
             ||||||||||||||||||||
```

```
Db       1966 TTTATTAAAAGTATCCAGC 1984
```

```
RESULT 2
AK032980
LOCUS       AK032980               1996 bp    mRNA    linear   HTC 18-SEP-2003
DEFINITION  Mus musculus 12 days embryo male wolffian duct includes surrounding
            region cDNA, RIKEN full-length enriched library, clone:6720484A10
            product:HEMATOPOIETIC PROGENITOR PROTEIN, full insert sequence.
ACCESSION   AK032980
VERSION     AK032980.1  GI:26328740
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 1996)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
```

```
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES              Location/Qualifiers
     source           1. .1996
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /strain="C57BL/6J"
                      /db_xref="FANTOM_DB:6720484A10"
                      /db_xref="MGI:2396648"
                      /db_xref="taxon:10090"
                      /clone="6720484A10"
                      /sex="male"
                      /tissue_type="wolffian duct includes surrounding region"
                      /clone_lib="RIKEN full-length enriched mouse cDNA library"
                      /dev_stage="12 days embryo"
     CDS              126. .839
                      /note="unnamed protein product; HEMATOPOIETIC PROGENITOR
                      PROTEIN (SPTR|Q99MP6, evidence: FASTY, 100%ID, 100%length,
                      match=711)
                      putative"
                      /codon_start=1
                      /protein_id="BAC28109.1"
                      /db_xref="GI:26328741"
                      /translation="MFARGLKRKYGDQEEGVEGFGTVPSYSLQRQSLLDMSLVKLQLC
                      HMLVEPNLCRSVLIANTVRQIQEEMSQDGVWHGMAPQNVDRAPVERLVSTEILCRTVR
                      GAEEEHPAPELEDAPLQNSVSELPIVGSAPGQRNPQSSLWEMDSPQENRGSFQKSLDQ
                      IFETLENKNSSSVEELFSDVDSSYYDLDTVLTGMMSGTKSSLCNGLEGFAAATPPPSS
                      TCKSDLAELDHVVEILVET"
     polyA_signal     1980. .1985
                      /note="putative"
     polyA_site       1996
                      /note="putative"
ORIGIN

  Query Match             90.5%;  Score 1883.8;  DB 11;  Length 1996;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1878;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        182 CTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGAAGGAGTA 241
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        117 CCGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGAAGGAGTA 176

Qy        242 GAGGGTTTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACAGTCACTCCTGGACATGTCC 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        177 GAGGGTTTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACAGTCACTCCTGGACATGTCC 236

Qy        302 CTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAATCTCTGCCGCTCGGTCCTC 361
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        237 CTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAATCTCTGCCGCTCGGTCCTC 296

Qy        362 ATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCAGGATGGTGTGTGGCATGGG 421
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        297 ATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCAGGATGGTGTGTGGCATGGG 356

Qy        422 ATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCTGGTGTCCACAGAGATCCTG 481
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        357 ATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCTGGTGTCCACAGAGATCCTG 416

Qy        482 TGTCGTACAGTGAGGGGAGCTGAGGAAGAGCACCCTGCTCCTGAACTGGAAGATGCTCCC 541
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        417 TGTCGTACAGTGAGGGGAGCTGAGGAAGAGCACCCTGCTCCTGAACTGGAAGATGCTCCC 476

Qy        542 TTGCAAAACTCGGTTTCCGAGCTCCCCATCGTTGGCTCAGCACCAGGGCAAAGGAACCCT 601
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        477 TTGCAAAACTCGGTTTCCGAGCTCCCCATCGTTGGCTCAGCACCAGGGCAAAGGAACCCT 536

Qy        602 CAGAGCAGCCTCTGGGAGATGGACAGCCCACAAGAAAACAGGGGAAGCTTTCAGAAGTCA 661
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        537 CAGAGCAGCCTCTGGGAGATGGACAGCCCACAAGAAAACAGGGGAAGCTTTCAGAAGTCA 596

Qy        662 CTGGACCAGATATTTGAGACCCTGGAGAACAAAAACTCCAGTTCAGTGGAGGAACTCTTC 721
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        597 CTGGACCAGATATTTGAGACCCTGGAGAACAAAAACTCCAGTTCAGTGGAGGAACTCTTC 656

Qy        722 TCAGATGTGGACAGCTCCTACTATGACCTGGACACAGTGCTAACAGGAATGATGAGTGGG 781
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        657 TCAGATGTGGACAGCTCCTACTATGACCTGGACACAGTGCTAACAGGAATGATGAGTGGG 716

Qy        782 ACCAAGTCCAGTCTCTGCAATGGCCTTGAGGGCTTTGCTGCAGCCACCCCTCCTCCCAGT 841
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        717 ACCAAGTCCAGTCTCTGCAATGGCCTTGAGGGCTTTGCTGCAGCCACCCCTCCTCCCAGT 776

Qy        842 TCCACTTGCAAGTCTGACCTGGCTGAGCTGGACCATGTGGTAGAGATTCTGGTGGAGACC 901
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        777 TCCACTTGCAAGTCTGACCTGGCTGAGCTGGACCATGTGGTAGAGATTCTGGTGGAGACC 836

Qy        902 TGAGAGGCCACCCCAGTGGGCTAAGGGTGAGGCCACCAGTCCCCATGGAGCTCACGTGTG 961
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        837 TGAGAGGCCACCCCAGTGGGCTAAGGGTGAGGCCACCAGTCCCCATGGAGCTCACGTGTG 896

Qy        962 TTGTGACCCAGAGACAGATAAGCACTTGTCCTAAGAGGGGCTCTGGCTCTTGAGCTCATT 1021
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        897 TTGTGACCCAGAGACAGATAAGCACTTGTCCTAAGAGGGGCTCTGGCTCTTGAGCTCATT 956

Qy       1022 ATCCTTTTGTGTGACATTGGACTCACTGTGGAGGATGGTGTGTCACAGCTATGTCTAGTC 1081
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        957 ATCCTTTTGTGTGACATTGGACTCACTGTGGAGGATGGTGTGTCACAGCTATGTCTAGTC 1016

Qy       1082 TATTTTCAATTAGATAGGTGAACTTTCTAAAATTAAGTTTTATATGTTTTTGGGCAATAT 1141
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1017 TATTTTCAATTAGATAGGTGAACTTTCTAAAATTAAGTTTTATATGTTTTTGGGCAATAT 1076

Qy       1142 TTTGTCTTAAGATATATTTTTTAAACTTTTTATACTTTAGATTTTTTTCAGCTATTTTCT 1201
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1077 TTTGTCTTAAGATATATTTTTTAAACTTTTTATACTTTAGATTTTTTTCAGCTATTTTCT 1136

Qy       1202 TAAAAGTATATTTTTTCTACAAACATCCTCTGCTGCTACATTAGAAACATTTATAACCTA 1261
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1137 TAAAAGTATATTTTTTCTACAAACATCCTCTGCTGCTACATTAGAAACATTTATAACCTA 1196

Qy       1262 AATACGATTGGTGTGTCATTTTAAAGGTTTAAATAGAAAACTTCTTTTGTTACTGAGTCT 1321
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1197 AATACGATTGGTGTGTCATTTTAAAGGTTTAAATAGAAAACTTCTTTTGTTACTGAGTCT 1256

Qy       1322 CTACACTCCCAAGGCAACTGTAAATGTAGCCGGCCGGGTGTTTACATGAGAGGCTCCAGT 1381
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1257 CTACACTCCCAAGGCAACTGTAAATGTAGCCGGCCGGGTGTTTACATGAGAGGCTCCAGT 1316

Qy       1382 ATGGTCTACATTCTAGTAGAGCTTGAAAAGAACCATGCACAGCTCCACTGCCCCCTCACT 1441
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1317 ATGGTCTACATTCTAGTAGAGCTTGAAAAGAACCATGCACAGCTCCACTGCCCCCTCACT 1376

Qy       1442 GGGTCTGCTCTGGCGGATCGGAGCTCTCTTCCTAGCCCCGTGTGCAGGATGGCTTTATTT 1501
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1377 GGGTCTGCTCTGGCGGATCGGAGCTCTCTTCCTAGCCCCGTGTGCAGGATGGCTTTATTT 1436

Qy       1502 ATGCCTATTTATATGTAAATGCCACTGAAAGCTAAGGTCTTACTCCTGGAAATCCCAACA 1561
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1437 ATGCCTATTTATATGTAAATGCCACTGAAAGCTAAGGTCTTACTCCTGGAAATCCCAACA 1496

Qy       1562 CCAGTTCTTCAGGGACTGCTGTGAGGCAGTGCCTTATGCAGGTCTTGTCCTTGGCCATCA 1621
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1497 CCAGTTCTTCAGGGACTGCTGTGAGGCAGTGCCTTATGCAGGTCTTGTCCTTGGCCATCA 1556

Qy       1622 CTGTCTGGTTCCCAGCCCAGCACATGTGACATGAGGACATGACATGCCCGAACCACCCAG 1681
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1557 CTGTCTGGTTCCCAGCCCAGCACATGTGACATGAGGACATGACATGCCCGAACCACCCAG 1616
```

```
Qy       1682 CACCACATGCTCCATGTCAAGTGTGTACGTGGAGACCACTGGCTCCCAGGCCTGTGCTCA 1741
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1617 CACCACATGCTCCATGTCAAGTGTGTACGTGGAGACCACTGGCTCCCAGGCCTGTGCTCA 1676

Qy       1742 GAGAGGGTGTGCAGTCCTACGTGTGCTGGGGGGGACGACGGTGACCTGTGCTTGCTTGCT 1801
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1677 GAGAGGGTGTGCAGTCCTACGTGTGCTGGGGGGGACGACGGTGACCTGTGCTTGCTTGCT 1736

Qy       1802 TTTAAAATGGTGCTTGGACGTTTTAAGGTTAAAAACAATCCGACTCCATATGATTTAGGG 1861
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1737 TTTAAAATGGTGCTTGGACGTTTTAAGGTTAAAAACAATCCGACTCCATATGATTTAGGG 1796

Qy       1862 CTCCTCCACCCTGGGGTGGCCCCTATGCTGTCTGCTTGGATCTCAAAGTCTTGGTACTCG 1921
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1797 CTCCTCCACCCTGGGGTGGCCCCTATGCTGTCTGCTTGGATCTCAAAGTCTTGGTACTCG 1856

Qy       1922 GCACTGTCAGACTCCACCCCATGTATCCTTTTTGTTTCTCTTGTGCTTTTTTTGGACTTC 1981
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1857 GCACTGTCAGACTCCACCCCATGTATCCTTTTTGTTTCTCTTGTGCTTTTTTTGGACTTC 1916

Qy       1982 CCAACCTGAGCCTAAGGTTTTATTTTATATGTGCTTCAATATCAACAATGTAAACCTCAC 2041
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1917 CCAACCTGAGCCTAAGGTTTTATTTTATATGTGCTTCAATATCAACAATGTAAACCTCAC 1976

Qy       2042 TTTATTAAAAGTATCCAGC 2060
             ||||||||||||||||||
Db       1977 TTTATTAAAAGTATCCAGC 1995
```

RESULT 3
BU511608

```
LOCUS       BU511608                1180 bp    mRNA    linear   EST 12-SEP-2002
DEFINITION  AGENCOURT_10129003 NIH_MGC_134 Mus musculus cDNA clone
            IMAGE:6506412 5', mRNA sequence.
ACCESSION   BU511608
VERSION     BU511608.1  GI:22817841
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1180)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. David Rowe
             cDNA Library Preparation: Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM14068  row: n  column: 13
            High quality sequence start: 12
            High quality sequence stop: 719.
FEATURES             Location/Qualifiers
     source          1. .1180
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6506412"
                     /tissue_type="undifferentiated limb"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_134"
                     /note="Vector: pCMV-SPORT6.1; Site_1: EcoRV; Site_2: NotI;
                     Cloned unidirectionally. Primer: Oligo dT. Average insert
                     size 1.7 kb. Constructed by ResGen, Invitrogen Corp. Note:
                     this is a NIH_MGC Library."
ORIGIN
```

Query Match 42.8%; Score 890.2; DB 13; Length 1180;
Best Local Similarity 89.4%; Pred. No. 3.5e-167;
Matches 1004; Conservative 0; Mismatches 107; Indels 12; Gaps 4;

```
Qy         98 GCGCCATGGCGACCCTTTGCTGGGCAGGTGACCGATTCCGGGTGCCCGAAGGAGCTGGCG 157
              |||||||||||||||||||  ||    |    |  ||||||||||||||||||    |
Db         42 GCGCCATGGCGACCCTTTGTGGG---CGCTGACCGATCCGGGTGCCCGAAGGAG--TGGC 96

Qy        158 TGGGTCTGCCTTGCAGCCGCCCGCCTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAA 217
              |||||||||||| ||       | ||||||||||||||||||||||||||||||||||
Db         97 GGGGTCTGCCTTGCGGC------GCCGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAA 150

Qy        218 TATGGTGACCAGGAAGAAGGAGTAGAGGGTTTTGGCACTGTCCCTTCCTATAGCCTGCAG 277
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        151 TATGGTGACCAGGAAGAAGGAGTAGAGGGTTTTGGCACTGTCCCTTCCTATAGCCTGCAG 210

Qy        278 CGACAGTCACTCCTGGACATGTCCCTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAG 337
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        211 CGACAGTCACTCCTGGACATGTCCCTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAG 270

Qy        338 CCCAATCTCTGCCGCTCGGTCCTCATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATG 397
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        271 CCCAATCTCTGCCGCTCGGTCCTCATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATG 330

Qy        398 AGCCAGGATGGTGTGTGGCATGGGATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAA 457
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        331 AGCCAGGATGGTGTGTGGCATGGGATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAG 390

Qy        458 CGCCTGGTGTCCACAGAGATCCTGTGTCGTACAGTGAGGGGAGCTGAGGAAGAGCACCCT 517
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        391 CGCCTGGTGTCCACAGAGATCCTGTGTCGTACAGTGAGGGGAGCTGAGGAAGAGCACCCT 450

Qy        518 GCTCCTGAACTGGAAGATGCTCCCTTGCAAAACTCGGTTTCCGAGCTCCCCATCGTTGGC 577
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        451 GCTCCTGAACTGGAAGATGCTCCCTTGCAAAACTCGGTTTCCGAGCTCCCCATCGTTGGC 510

Qy        578 TCAGCACCAGGGCAAAGGAACCCTCAGAGCAGCCTCTGGGAGATGGACAGCCCACAAGAA 637
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        511 TCAGCACCAGGGCAAAGGAACCCTCAGAGCAGCCTCTGGGAGATGGACAGCCCACAAGAA 570

Qy        638 AACAGGGGAAGCTTTCAGAAGTCACTGGACCAGATATTTGAGACCCTGGAGAACAAAAAC 697
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        571 AACAGGGGAAGCTTTCAGAAGTCACTGGACCAGATATTTGAGACCCTGGAGAACAAAAAC 630

Qy        698 TCCAGTTCAGTGGAGGAACTCTTCTCAGATGTGGACAGCTCCTACTATGACCTGGACACA 757
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        631 TCCAGTTCAGTGGAGGAACTCTTCTCAGATGTGGACAGCTCCTACTATGACCTGNACACA 690

Qy        758 GTGCTAACAGGAATGATGAGTGGGACCAAGTCCAGTCTCTGCAATGGCCTTGAGGGCTTT 817
              |||||||||||||||||| ||||||| |||||||||||||  ||||||||||||||||||
Db        691 GTGCTAACAGGAATGATGACTGGGACCNAGTCCAGTCTCTGNCATGGCCTTGAGGGCTTT 750

Qy        818 GCTGCAGCCACCCCTCCTCCCAGTTCCACTTGCAAGTCTGACCTGGCTGAGCTGGACCAT 877
              |||||||| |||||||||||||||||||    |||||||||| ||||||||||||||||||
Db        751 GCTGCAGCCGCCCCTCCTCCCAGTTCCACTTTGCAGTCTGACCTGNCTGAGCTGGACCAT 810

Qy        878 GTGGTAGAGATTCTGGTGGAGACCTGAGAGGCCACCCCAGTGGGCTAAGGGTGAGGCCAC 937
              |||||||||||||||||||||||||||||| |||||||| |||||||||||||| |||
Db        811 GTGGTAGAGATTCTGGTGGAGACCTGAGAAGCCACCCCCGTGGGCTAAGGGTGAAGCCCA 870

Qy        938 CAGTCCCCATGGAGCTCACGTGTGTTGTGACCCAGAGACAGATAAGCACTTGTCCTAAGA 997
              |||||||||||||||||||||||| |||||||| |||||||||| |||||| |||| |
Db        871 CAGTCCCCATGGAGCTCACGTGTGGTGTGACCCCGAGACAGATAAACACTTGGTCTAAAA 930

Qy        998 GGGGCTCTGGCTCTTGAGCTCATTATCCTTTTGTGTGACATTGGACTCACTGTGGAGGAT 1057
              |||| || || |||||  ||| |||||||||| | |||||| |    |  | | | |||
Db        931 GGGGTTCCGGGTCTTGGACTCCTTATCCTTTTGGGGGACATGGACTCCCTGGGGAGGGAG 990

Qy       1058 GGTGTGTCACAGCTATGTCTAGTCTATTTTCAATTAGATAGGTGAACTTTCTAAAATTAA 1117
              || | ||| || ||||| |  | |||||| |||| | ||| ||  || | ||| | |
Db        991 GGGGGGTCCCACCTATGCTCAACCAATTTTCCATTAAAAAGGGGACTTTCCAAAATTAAG 1050

Qy       1118 GTTTTATATGTTTTTGGGCAATA-TTTTGTCTTAAGATATATTTTTTAAACTTTTTATAC 1176
              |||    || |||||||||| | |||||||||||| | |||||||||||||||||||| |
```

```
Db       1051 TTTTATATTGGTTTTGGGCAAAATTTTTGTCTTAAAAAATATTTTTTAAACTTTTTAAAA 1110

Qy       1177 TTTAGATTTTTTTCAGCTATTTTCTTAAAAGTATATTTTTTCT 1219
              ||| ||||||||     ||||| |   ||  | | ||||| |
Db       1111 CTTAAATTTTTTTTTCGCATTTTTTCCTAAAAAAAATTTTTTT 1153


RESULT 4
AV000968
LOCUS       AV000968                989 bp    mRNA    linear   EST 05-SEP-2001
DEFINITION  AV000968 Mus musculus C57BL/6J kidney Mus musculus cDNA clone
            0610010F24, mRNA sequence.
ACCESSION   AV000968
VERSION     AV000968.2  GI:15431363
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 989)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T.,
            Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J.,
            Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K.,
            Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     On May 11, 1999 this sequence version replaced gi:4777818.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
            and Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
            Sugahara,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Yamanaka,I., Kiyosawa,H., Kondo,S., Saito,T., Shinagawa,A.,
            Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K.,
            Arakawa,T., Ishii,Y. and Hayashizaki,Y.
             Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct.
            Func. Genomics 2 pre, L72-L86 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
FEATURES              Location/Qualifiers
     source           1. .989
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="0610010F24"
                      /tissue_type="kidney"
                      /clone_lib="Mus musculus C57BL/6J kidney"
ORIGIN

  Query Match             40.8%;  Score 849.2;  DB 9;  Length 989;
  Best Local Similarity   92.9%;  Pred. No. 5.5e-159;
  Matches  890;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy       1103 ACTTTCTAAAATTAAGTTTTATATGTTTTTGGGCAATATTTTGTCTTAAGATATATTTTT 1162
              | ||  | | |     |  || ||  ||  |     || | ||| | | |  | | |
Db         31 ATTTCATTAGAAGGGGACTTCTAAATTAGTATTATATGTGTTGGCATTATTTGTCTAAGA 90

Qy       1163 TAAACTTTTTATACTTTAGATTTTTTTCAGCTATTTTCTTAAAAGTATATTTTTTCTACA 1222
              |||  |||  |   | ||  ||   ||   |   | |||   ||       ||||||||||
Db         91 TAATTTTTAAACTTTATACTTTAGATTTTTCAGCTATCTCTTAAAGTATATTTTTCTACA 150

Qy       1223 AACATCCTCTGCTGCTACATTAGAAACATTTATAACCTAAATACGATTGGTGTGTCATTT 1282
              |||| |||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        151 AACACCCTCTGCTGCTACATTAGAAACATCTATAACCTAAATACGATTGGTGTGTCATTA 210

Qy       1283 TAAAGGTTTAAATAGAAAACTTCTTTTGTTACTGAGTCTCTACACTCCCAAGGCAACTGT 1342
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        211 TAAAGGTTTAAATAGAAAACTTCTTTTGTTACTGAGTCTCTACACTCCCAAGGCAACTGT 270

Qy       1343 AAATGTAGCCGGCCGGGTGTTTACATGAGAGGCTCCAGTATGGTCTACATTCTAGTAGAG 1402
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        271 AAATGTAGCCGGCCGGGTGTTTACATGAGAGGCTCCAGTATGGTCTACATTCTAGTAGAG 330

Qy       1403 CTTGAAAAGAACCATGCACAGCTCCACTGCCCCCTCACTGGGTCTGCTCTGGCGGATCGG 1462
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        331 CTTGAAAAGAACCATGCACAGCTCCACTGCCCCCTCACTGGGTCTGCTCTGGCGGATCGG 390

Qy       1463 AGCTCTCTTCCTAGCCCCGTGTGCAGGATGGCTTTATTTATGCCTATTTATATGTAAATG 1522
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        391 AGCTCTCTTCCTAGCCCCGTGTGCAGGATGGCTTTATTTATGCCTATTTATATGTAAATG 450

Qy       1523 CCACTGAAAGCTAAGGTCTTACTCCTGGAAATCCCAACACCAGTTCTTCAGGGACTGCTG 1582
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        451 CCACTGAAAGCTAAGGTCTTACTCCTGGAAATCCCAACACCAGTTCTTCAGGGACTGCTG 510

Qy       1583 TGAGGCAGTGCCTTATGCAGGTCTTGTCCTTGGCCATCACTGTCTGGTTCCCAGCCCAGC 1642
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        511 TGAGGCAGTGCCTTATGCAGGTCTTGTCCTTGGCCATCACTGTCTGGTTCCCAGCCCAGC 570

Qy       1643 ACATGTGACATGAGGACATGACATGCCCGAACCACCCAGCACCACATGCTCCATGTCAAG 1702
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        571 ACATGTGACATGAGGACATGACATGCCCGAACCACCCAGCACCACATGCTCCATGTCAAG 630

Qy       1703 TGTGTACGTGGAGACCACTGGCTCCCAGGCCTGTGCTCAGAGAGGGTGTGCAGTCCTACG 1762
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        631 TGTGTACGTGGAGACCACTGGCTCCCAGGCCTGTGCTCAGAGAGGGTGTGCAGTCCTACG 690

Qy       1763 TGTGCTGGGGGGGACGACGGTGACCTGTGCTTGCTTGCTTTTAAAATGGTGCTTGGACGT 1822
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        691 TGTGCTGGGGGGGACGACGGTGACCTGTGCTTGCTTGCTTTTAAAATGGTGCTTGGACGT 750

Qy       1823 TTTAAGGTTAAAAACAATCCGACTCCATATGATTTAGGGCTCCTCCACCCTGGGGTGGCC 1882
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        751 TTTAAGGTTAAAAACAATCCGACTCCATATGATTTAGGGCTCCTCCACCCGGGGGTGGCC 810

Qy       1883 CCTATGCTGTCTGCTTGGATCTCAAAGTCTTGGTACTCGGCACTGTCAGACTCCACCCCA 1942
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        811 CCTATGCTGTCTGCTTGGGTCTCAAAGTCTTGGTACTCGGCACTGTCAGACTCCACCCCA 870

Qy       1943 TGTATCCTTTTTGTTTCTCTTGTGCTTTTTTTGGACTTCCCAACCTGAGCCTAAGGTTTT 2002
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        871 TGTATCCTTTTTGTTTCTCTTGTGCTTTTTTTGGACTTCCCAACCTGAGCCTAAGGTTTT 930

Qy       2003 ATTTTATATGTGCTTCAATATCAACAATGTAAACCTCACTTTATTAAAAGTATCCAGC 2060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        931 ATTTTATATGTGCTTCAATATCAACAATGTAAACCTCACTTTATTAAAAGTATCCAGC 988
```

```
RESULT 5
BU054465
LOCUS       BU054465                836 bp    mRNA    linear   EST 26-AUG-2002
DEFINITION  UI-M-FD0-bzj-e-04-0-UI.r1 NIH_BMAP_FD0 Mus musculus cDNA clone
            IMAGE:6404331 5', mRNA sequence.
ACCESSION   BU054465
VERSION     BU054465.1  GI:22494529
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 836)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. James Lin, Univeristy of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1. .836
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6404331"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 12.5 dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_FD0"
                     /note="Organ: brain; Vector: pYX-Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according to
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured mRNa was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with an
                     oligo-dT primer containing a Not I site. Double stranded
                     cDNA was size selected according to mRNA size fraction,
                     ligated with EcoR I adaptor, digested with Not I, and then
                     cloned directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA
                     tail, is TGAGAGAGCC. This library was created for the
                     University of Iowa Mouse Brain Molecular Anatomy Project
                     (BMAP): 'Gene Discovery in the Developing Mouse Nervous
                     System', supported by National Instututes of Mental Health
                     (NIMH), Hemin Chin, Ph.D., program coordinator."
ORIGIN

  Query Match           38.9%;  Score 810.4;  DB 13;  Length 836;
  Best Local Similarity   99.5%;  Pred. No. 3.2e-151;
  Matches  833;  Conservative    0;  Mismatches    2;  Indels    2;  Gaps    2;

Qy        252 GCACTGTCCCTTCCTATAGCCTGCAGCGACAGTCACTCCT-GGACATGTCCCTTGTCAAG 310
              ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db          1 GCACTGTCCCTTCCTATAGCCTGCAGCGACAGTCACTCCTAGGACATGTCCCTTGTCAAG 60

Qy        311 CTCCAGCTCTGTCACATGCTAGTGGAGCCCAATCTCTGCCGCTCGGTCCTCATCGCCAAC 370
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CTCCAGCTCTGTCACATGCTAGTGGAGCCCAATCTCTGCCGCTCGGTCCTCATCGCCAAC 120

Qy        371 ACAGTCCGGCAGATCCAGGAGGAAATGAGCCAGGATGGTGTGTGGCATGGGATGGCACCC 430
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACAGTCCGGCAGATCCAGGAGGAAATGAGCCAGGATGGTGTGTGGCATGGGATGGCACCC 180

Qy        431 CAGAATGTAGATCGGGCACCAGTTGAACGCCTGGTGTCCACAGAGATCCTGTGTCGTACA 490
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGAATGTAGATCGGGCACCAGTTGAACGCCTGGTGTCCACAGAGATCCTGTGTCGTACA 240

Qy        491 GTGAGGGGAGCTGAGGAAGAGCACCCTGCTCCTGAACTGGAAGATGCTCCCCTTGCAAAAC 550
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTGAGGGGAGCTGAGGAAGAGCACCCTGCTCCTGAACTGGAAGATGCTCCCCTTGCAAAAC 300

Qy        551 TCGGTTTCCGAGCTCCCCATCGTTGGCTCAGCACCAGGGCAAAGGAACCCTCAGAGCAGC 610
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCGGTTTCCGAGCTCCCCATCGTTGGCTCAGCACCAGGGCAAAGGAACCCTCAGAGCAGC 360

Qy        611 CTCTGGGAGATGGACAGCCCACAAGAAAACAGGGGAAGCTTTCAGAAGTCACTGGACCAG 670
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTCTGGGAGATGGACAGCCCACAAGAAAACAGGGGAAGCTTTCAGAAGTCACTGGACCAG 420

Qy        671 ATATTTGAGACCCTGGAGAACAAAAACTCCAGTTCAGTGGAGGAACTCTTCTCAGATGTG 730
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        421 ATATTTGAGACCCTGGAGAAC-AAAACTCCAGTTCAGTGGAGGAACTCTTCTCAGATGTG 479

Qy        731 GACAGCTCCTACTATGACCTGGACACAGTGCTAACAGGAATGATGAGTGGGACCAAGTCC 790
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        480 GACAGCTCCTACTATGACCTGGACACAGTGCTAACAGGAATGATGAGTGGGACCAAGTCC 539

Qy        791 AGTCTCTGCAATGGCCTTGAGGGCTTTGCTGCAGCCACCCCTCCTCCCAGTTCCACTTGC 850
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        540 AGTCTCTGCAATGGCCTTGAGGGCTTTGCTGCAGCCACCCCTCCTCCCAGTTCCACTTGC 599

Qy        851 AAGTCTGACCTGGCTGAGCTGGACCATGTGGTAGAGATTCTGGTGGAGACCTGAGAGGCC 910
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        600 AAGTCTGACCTGGCTGAGCTGGACCATGTGGTAGAGATTCTGGTGGAGACCTGAGAGGCC 659

Qy        911 ACCCCAGTGGGCTAAGGGTGAGGCCACCAGTCCCCATGGAGCTCACGTGTGTTGTGACCC 970
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        660 ACCCCAGTGGGCTAAGGGTGAGGCCACCAGTCCCCATGGAGCTCACGTGTGTTGTGANCC 719

Qy        971 AGAGACAGATAAGCACTTGTCCTAAGAGGGGCTCTGGCTCTTGAGCTCATTATCCTTTTG 1030
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        720 AGAGACAGATAAGCACTTGTCCTAAGAGGGGCTCTGGCTCTTGAGCTCAATATCCTTTTG 779

Qy       1031 TGTGACATTGGACTCACTGTGGAGGATGGTGTGTCACAGCTATGTCTAGTCTATTTT 1087
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        780 TGTGACATTGGACTCACTGTGGAGGATGGTGTGTCACAGCTATGTCTAGTCTATTTT 836


RESULT 6
CB183964
LOCUS       CB183964                919 bp    mRNA    linear   EST 31-JAN-2003
DEFINITION  AGENCOURT_11380446 NIH_MGC_164 Mus musculus cDNA clone
            IMAGE:30245029 5', mRNA sequence.
ACCESSION   CB183964
VERSION     CB183964.1  GI:28183220
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 919)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. David Rowe and Dr. Mina
             cDNA Library Preparation: Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDAM0318  row: n  column: 14
            High quality sequence stop: 674.
```

```
FEATURES             Location/Qualifiers
     source          1. .919
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:30245029"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_164"
                     /note="Vector: pCMV-SPORT6.1; Site_1: EcoRV; Site_2: NotI;
                     Non-normalized full-length enriched library from pooled
                     mouse embryonic limb, maxilla and mandible, day 10.5 and
                     11.5 (size selected for the 0.5-1 kb fragments) Cloned
                     directionally, priming method: Oligo-dT.  cDNA enrichment:
                     >1k bp, Average insert size 1.8k bp. Priming sequence:
                     5'GACTAGTTCTAGATCGCGAGCGGCCGCCC(T) 3'. Tissue contributed
                     by, David Rowe. Library constructed by ResGen, Invitrogen
                     Corp."
ORIGIN

  Query Match           38.9%;  Score 809.8;  DB 14;  Length 919;
  Best Local Similarity   95.5%;  Pred. No. 4e-151;
  Matches  878;  Conservative    0;  Mismatches   32;  Indels    9;  Gaps    4;

Qy       1064 TCACAGCTATGTCTAGTCTATTTTCAATTAGATAGGTGAACTTTCTAAAATTAAGTTTTA 1123
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TCACAGCTATGTCTAGTCTATTTTCAATTAGATAGGTGAACTTTCTAAAATTAAGTTTTA 60

Qy       1124 TATGTTTTTGGGCAATATTTTGTCTTAAGATATATTTTTTAAACTTTTTATACTTTAGAT 1183
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TATGTTTTTGGGCAATATTTTGTCTTAAGATATATTTTTTAAACTTTTTATACTTTAGAT 120

Qy       1184 TTTTTTCAGCTATTTTCTTAAAAGTATATTTTTTCTACAAACATCCTCTGCTGCTACATT 1243
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTTTTTCAGCTATTTTCTTAAAAGTATATTTTTTCTACAAACATCCTCTGCTGCTACATT 180

Qy       1244 AGAAACATTTATAACCTAAATACGATTGGTGTGTCATTTTAAAGGTTTAAATAGAAAACT 1303
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGAAACATTTATAACCTAAATACGATTGGTGTGTCATTTTAAAGGTTTAAATAGAAAACT 240

Qy       1304 TCTTTTGTTACTGAGTCTCTACACTCCCAAGGCAACTGTAAATGTAGCCGGCCGGGTGTT 1363
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCTTTTGTTACTGAGTCTCTACACTCCCAAGGCAACTGTAAATGTAGCCGGCCGGGTGTT 300

Qy       1364 TACATGAGAGGCTCCAGTATGGTCTACATTCTAGTAGAGCTTGAAA--AGAACCATGCAC 1421
             ||||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||
Db        301 TACATGAGAGGCTCCAGTATGGTCTACATTCTAGTAGAGCTTGAAAAGAGAACCATGCAC 360

Qy       1422 AGCTCCACTGCCCCCTCACTGGGTCTGCTCTGGCGGATCGGAGCTCTCTTCCTAGCCCCG 1481
             ||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGCTCCACTGCCCCCTCATTGGGTCTGCTCTGGCGGATCGGAGCTCTCTTCCTAGCCCCG 420

Qy       1482 TGTGCAGGATGGCTTTATTTATGCCTATTTATATGTAAATGCCACTGAAAGCTAAGGTCT 1541
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TGTGCAGGATGGCTTTATTTATGCCTATTTATATGTAAATGCCACTGAAAGCTAAGGTCT 480

Qy       1542 TACTCCTGGAAATCCCAACACCAGTTCTTCAGGGACTGCTGTGAGGCAGTGCCTTATGCA 1601
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TACTCCTGGAAATCCCAACACCAGTTCTTCAGGGACTGCTGTGAGGCAGTGCCTTATGCA 540

Qy       1602 GGTCTTG----TCCTTGGCCATCACTGTCTGGTTCCCAGCCCAGCACATGTGACATGAGG 1657
             |||||||    ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGTCTTGCTTGTCCTTGGCCATCACTGTCTGGTTCCCAGCCCAGCACATGTGACATGAGG 600

Qy       1658 ACATGACATGCCCGAACCACCCAGCACCACATGCTCCATGTCAAGTGTGTACGTGGAGAC 1717
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ACATGACATGCCCGAACCACCCAGCACCACATGCTCCATGTCAAGTGTGTACGTGGAGAC 660

Qy       1718 CACTGGCTCCCAGGCCTGTGCTCAGAGAGGGTGTGCAGTCCTACGTGTGCTGGGGGGGAC 1777
             |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        661 CACTGGCTCCCAGGCCTGTGCTCAGAGAGGGTGTGCAGTCCCACGTGTGCTGGGGGGGAC 720

Qy       1778 GACGGTGACCTGTGCTTGCTTGCTTTTAAAATGGTGCTTGGACGTTTTAAGGTTAAAAAC 1837
             |||||||||||||||||||||||||| |||||||||||||| ||||||||||||||||||
Db        721 GACGGTGACCTGTGCTTGCTTGCTTTT-AAATGGTGCTTGGATGTTTTAAGGTTAAAAAC 779

Qy       1838 AATCCGACTCCATATGATTTAGGGCTCCTCCACCCTGGGGTGGCCCCCTATGCTGTCTGCT 1897
             |||||||||||||||||||||||||||||||||||||  |||  |||||||||||| |||||||
Db        780 AATCCGACTCCATATGATTTAGGGCTCCTCCACCTGGGGTGGGCCCCTATGCGGTCTGCT 839

Qy       1898 TGGATCTCAAAGTCTTGGTACTCGG--CACTGTCAGACTCCACCCCATGTATCCTTTTTG 1955
             |||||||||||||||||||||||||      |||  |||| |||||  | | ||  |||
Db        840 TGGATCTCAAAGTCTTGGTACTCGGGCACTGGTCGAACTCTACCCCCAGAAACCCCTTTT 899

Qy       1956 TTTCTCTTGTGCTTTTTTT 1974
              | ||   ||    ||||
Db        900 GGTTTCCCCTGGGGCTTTT 918


RESULT 7
BQ950078
LOCUS       BQ950078                927 bp    mRNA    linear   EST 21-AUG-2002
DEFINITION  AGENCOURT_8888587 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6311335
            5', mRNA sequence.
ACCESSION   BQ950078
VERSION     BQ950078.1  GI:22365556
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 927)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Susan L. Sullivan, PhD.
             cDNA Library Preparation: ResGen, Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13735  row: b  column: 08
            High quality sequence stop: 608.
FEATURES             Location/Qualifiers
     source          1. .927
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6311335"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_129"
                     /note="Organ: olfactory epithelium; Vector: pCMV-SPORT6.1;
                     Site_1: EcoRV; Site_2: NotI; Cloned unidirectionally.
                     Primer: Oligo dT. Average insert size 2.2 kb. Constructed
                     by ResGen, Invitrogen Corp. Note: this is a NIH_MGC
                     Library."
ORIGIN

  Query Match           38.3%;  Score 798;  DB 13;  Length 927;
  Best Local Similarity   96.6%;  Pred. No. 9.1e-149;
  Matches  858;  Conservative    0;  Mismatches   25;  Indels    5;  Gaps    4;

Qy        182 CTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGAAGGAGTA 241
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         22 CCGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGAAGGAGTA 81

Qy        242 GAGGGTTTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACAGTCACTCCTGGACATGTCC 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         82 GAGGGTTTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACAGTCACTCCTGGACATGTCC 141

Qy        302 CTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAATCTCTGCCGCTCGGTCCTC 361
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        142 CTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAATCTCTGCCGCTCGGTCCTC 201
```

```
Qy        362 ATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCAGGATGGTGTGTGGCATGGG 421
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        202 ATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCAGGATGGTGTGTGGCATGGG 261

Qy        422 ATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCTGGTGTCCACAGAGATCCTG 481
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        262 ATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCTGGTGTCCACAGAGATCCTG 321

Qy        482 TGTCGTACAGTGAGGGGAGCTGAGGAAGAGCACCCTGCTCCTGAACTGGAAGATGCTCCC 541
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        322 TGTCGTACAGTGAGGGGAGCTGAGGAAGAGCACCCTGCTCCTGAACTGGAAGATGCTCCC 381

Qy        542 TTGCAAAACTCGGTTTCCGAGCTCCCCATCGTTGGCTCAGCACCAGGGCAAAGGAACCCT 601
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        382 TTGCAAAACTCGGTTTCCGAGCTCCCCATCGTTGGCTCAGCACCAGGGCAAAGGAACCCT 441

Qy        602 CAGAGCAGCCTCTGGGAGATGGACAGCCCACAAGAAAACAGGGGAAGCTTTCAGAAGTCA 661
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        442 CAGAGCAGCCTCTGGGAGATGGACAGCCCACAAGAAAACAGGGGAAGCTTTCAGAAGTCA 501

Qy        662 CTGGACCAGATATTTGAGACCCTGGAGAACAAAAACTCCAGTTCAGTGGAGGAACTCTTC 721
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        502 CTGGACCAGATATTTGAGACCCTGGAGAACAAAAACTCCAGTTCAGTGGAGGAACTCTTC 561

Qy        722 TCAGATGTGGACAGCTCCTACTATGACCTGGACACAGTGCTAACAGGAATGATGAGTGGG 781
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        562 TCAGATGTGGACAGCTCCTACTATGACCTGGACACAGTGCTAACAGGAATGATGAGTGGG 621

Qy        782 ACCAAGTCCAGTCTCTGCAATGGCCTTGAGGGCTTTGCTGCAGCCACCCCTCCTCCCAGT 841
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        622 ACCAAGTCCAGTCTCTGCAATGGCCTTGAGGGCTTTGCTGCAGCCACCCCTCCTCCCAGT 681

Qy        842 TCCACTTGCAAGTCTGACCTGGCTGAGCTGGACCATGTGGTAGAGATTCTGGTGGAGACC 901
              |||||||||||||||||||||||||||||||||||||||| |||||||| |||||||
Db        682 TCCACTTGCAAGTCTGACCTGGCTGAGCTGGACCATGTGGTAAAGATTCTGGGGGAGACC 741

Qy        902 TGAGAGGCCACCCCAGTGGG-CTAAGGGTGAGGCCACCAGTCCCCATGGAGCTCACGTGT 960
              ||||| |||||||||||||| |||||||||||||||||||| |||||||||||||||||||
Db        742 TGAGAAGCCACCCCAGTGGGCCTAAGGGTGAGGCCACCAGTCCCCCTGGAGCTCACGTGT 801

Qy        961 GTTGTGACCCAGAGACAGATAAGCACTT-GTCCTAAGAGGGGCTCTGG--CTCTTGAGCT 1017
              | |||||||||| |||||||||||||||| |||||||||||||||||    || |||||
Db        802 GGTGTGACCCAAAGACAGATAAGCACTTGGTCCTAAGAGGGGCTCTGGGCTCCTGGAGCT 861

Qy       1018 CATT-ATCCTTTTGTGTGACATTGGACTCACTGTGGAGGATGGTGTGT 1064
              |||| |||||||||||| |  |||||   | |   || | | || | ||
Db        862 CATTAATCCTTTTGTGGGGAATTGGGACCCCCTGGGGGAAAGGAGGGT 909


RESULT 8
BU148336
LOCUS       BU148336                920 bp    mRNA    linear   EST 03-SEP-2002
DEFINITION  AGENCOURT_8753844 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6393860
            5', mRNA sequence.
ACCESSION   BU148336
VERSION     BU148336.1  GI:22661868
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 920)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,
            Ph.D.
             cDNA Library Preparation: ResGen, Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM13885  row: p  column: 21
             High quality sequence stop: 740.
FEATURES             Location/Qualifiers
     source          1. .920
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6393860"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_130"
                     /note="Organ: otocysts; Vector: pCMV-SPORT6.1; Site_1:
                     EcoRV; Site_2: NotI; Cloned unidirectionally.  Primer:
                     Oligo dT.  Average insert size 1.95  kb.  Constructed by
                     ResGen, Invitrogen Corp. Note: this is a NIH_MGC Library."
ORIGIN

  Query Match             38.0%;  Score 790.6;  DB 13;  Length 920;
  Best Local Similarity   98.7%;  Pred. No. 2.7e-147;
  Matches  815;  Conservative    0;  Mismatches    9;  Indels    2;  Gaps    2;

Qy        182 CTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGAAGGAGTA 241
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         46 CCGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGAAGGAGTA 105

Qy        242 GAGGGTTTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACAGTCACTCCTGGACATGTCC 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        106 GAGGGTTTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACAGTCACTCCTGGACATGTCC 165

Qy        302 CTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAATCTCTGCCGCTCGGTCCTC 361
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        166 CTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAATCTCTGCCGCTCGGTCCTC 225

Qy        362 ATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCAGGATGGTGTGTGGCATGGG 421
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        226 ATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCAGGATGGTGTGTGGCATGGG 285

Qy        422 ATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCTGGTGTCCACAGAGATCCTG 481
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        286 ATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCTGGTGTCCACAGAGATCCTG 345

Qy        482 TGTCGTACAGTGAGGGGAGCTGAGGAAGAGCACCCTGCTCCTGAACTGGAAGATGCTCCC 541
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        346 TGTCGTACAGTGAGGGGAGCTGAGGAAGAGCACCCTGCTCCTGAACTGGAAGATGCTCCC 405

Qy        542 TTGCAAAACTCGGTTTCCGAGCTCCCCATCGTTGGCTCAGCACCAGGGCAAAGGAACCCT 601
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        406 TTGCAAAACTCGGTTTCCGAGCTCCCCATCGTTGGCTCAGCACCAGGGCAAAGGAACCCT 465

Qy        602 CAGAGCAGCCTCTGGGAGATGGACAGCCCACAAGAAAACAGGGGAAGCTTTCAGAAGTCA 661
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        466 CAGAGCAGCCTCTGGGAGATGGACAGCCCACAAGAAAACAGGGGAAGCTTTCAGAAGTCA 525

Qy        662 CTGGACCAGATATTTGAGACCCTGGAGAACAAAAACTCCAGTTCAGTGGAGGAACTCTTC 721
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        526 CTGGACCAGATATTTGAGACCCTGGAGAACAAAAACTCCAGTTCAGTGGAGGAACTCTTC 585

Qy        722 TCAGATGTGGACAGCTCCTACTATGACCTGGACACAGTGCTAACAGGAATGATGAGTGGG 781
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        586 TCAGATGTGGACAGCTCCTACTATGACCTGGACACAGTGCTAACAGGAATGATGAGTGGG 645

Qy        782 ACCAAGTCCAGTCTCTGCAATGGCCTTGAGGGCTTTGCTGCAGCCACCCCTCCTCCCAGT 841
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        646 ACCAAGTCCAGTCTCTGCAATGGCCTTGAGGGCTTTGCTGCAGCCACCCCTCCTCCCAGT 705

Qy        842 TCCACTTGCAAGTCTGACCTGGCTGAGCTGGACCATGTGGTAGAGATTCTGGTGGAGACC 901
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        706 TCCACTTGCAAGTCTGACCTGGCTGAGCTGGACCATGTGGTAGAGATTCTGGTGGAGACC 765

Qy        902 TGAGAGGCCACCCCAGTGGGCTAAGGGTGAGGCCACCAGTCCCCATGG-AGCTCACGTGT 960
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        766 TGAGAGGCCACCCCAGTGGGCTAAGGGTGAGGCCACCAGTCCCCATGGNAGCTCACGTGT 825
```

```
Qy        961 GTTGT-GACCCAGAGACAGATAAGCACTTGTCCTAAGAGGGGCTCT 1005
              | ||| |||||||||||||||||||||||||||  |   |||  ||
Db        826 GGTGTGGACCCAGAGACAGATAAGCACTTGTCCCTAAGAGGGNGCT 871


RESULT 9
BU151870
LOCUS       BU151870                926 bp    mRNA    linear   EST 03-SEP-2002
DEFINITION  AGENCOURT_8754199 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6332467
            5', mRNA sequence.
ACCESSION   BU151870
VERSION     BU151870.1  GI:22665402
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 926)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,
            Ph.D.
             cDNA Library Preparation: ResGen, Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13790  row: b  column: 20
            High quality sequence stop: 660.
FEATURES             Location/Qualifiers
     source          1. .926
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6332467"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_130"
                     /note="Organ: otocysts; Vector: pCMV-SPORT6.1; Site_1:
                     EcoRV; Site_2: NotI; Cloned unidirectionally.  Primer:
                     Oligo dT.  Average insert size 1.95  kb.  Constructed by
                     ResGen, Invitrogen Corp. Note: this is a NIH_MGC Library."
ORIGIN

  Query Match           37.7%;  Score 785.8;  DB 13;  Length 926;
  Best Local Similarity   95.8%;  Pred. No. 2.4e-146;
  Matches  883;  Conservative    0;  Mismatches   25;  Indels   14;  Gaps    7;

Qy        134 TCCGGGTGCCCGAAGGAGCTGGCGTGGGTCTGCCTTGCAGCCGCCCGCCTGGACAGGATG 193
              |||||||||||||||||  |   |||||||||||| ||       | ||||||||||
Db          6 TCCGGGTGCCCGAAGGAG--TGGCGGGGTCTGCCTTGCGGC------GCCGGACAGGATG 57

Qy        194 TTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGAAGGAGTAGAGGGTTTTGGC 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         58 TTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGAAGGAGTAGAGGGTTTTGGC 117

Qy        254 ACTGTCCCTTCCTATAGCCTGCAGCGACAGTCACTCCTGGACATGTCCCTTGTCAAGCTC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        118 ACTGTCCCTTCCTATAGCCTGCAGCGACAGTCACTCCTGGACATGTCCCTTGTCAAGCTC 177

Qy        314 CAGCTCTGTCACATGCTAGTGGAGCCCAATCTCTGCCGCTCGGTCCTCATCGCCAACACA 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        178 CAGCTCTGTCACATGCTAGTGGAGCCCAATCTCTGCCGCTCGGTCCTCATCGCCAACACA 237

Qy        374 GTCCGGCAGATCCAGGAGGAAATGAGCCAGGATGGTGTGTGGCATGGGATGGCACCCCAG 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        238 GTCCGGCAGATCCAGGAGGAAATGAGCCAGGATGGTGTGTGGCATGGGATGGCACCCCAG 297

Qy        434 AATGTAGATCGGGCACCAGTTGAACGCCTGGTGTCCACAGAGATCCTGTGTCGTACAGTG 493
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        298 AATGTAGATCGGGCACCAGTTGAACGCCTGGTGTCCACAGAGATCCTGTGTCGTACAGTG 357

Qy        494 AGGGGAGCTGAGGAAGAGCACCCTGCTCCTGAACTGGAAGATGCTCCCTTGCAAAACTCG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        358 AGGGGAGCTGAGGAAGAGCACCCTGCTCCTGAACTGGAAGATGCTCCCTTGCAAAACTCG 417

Qy        554 GTTTCCGAGCTCCCCATCGTTGGCTCAGCACCAGGGCAAAGGAACCCTCAGAGCAGCCTC 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        418 GTTTCCGAGCTCCCCATCGTTGGCTCAGCACCAGGGCAAAGGAACCCTCAGAGCAGCCTC 477

Qy        614 TGGGAGATGGACAGCCCACAAGAAAACAGGGGAAGCTTTCAGAAGTCACTGGACCAGATA 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        478 TGGGAGATGGACAGCCCACAAGAAAACAGGGGAAGCTTTCAGAAGTCACTGGACCAGATA 537

Qy        674 TTTGAGACCCTGGAGAACAAAAACTCCAGTTCAGTGGAGGAACTCTTCTCAGATGTGGAC 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        538 TTTGAGACCCTGGAGAACAAAAACTCCAGTTCAGTGGAGGAACTCTTCTCAGATGTGGAC 597

Qy        734 AGCTCCTACTATGACCTGGACACAGTGCTAACAGGAATGATGAGTGGGACCAAGTCCAGT 793
              |||||||||||||||||| ||||||||||||||||||||||||||| ||||||||||||
Db        598 AGCTCCTACTATGACCTGNACACAGTGCTAACAGGAATGATGAGTGNGACCAAGTCCAGT 657

Qy        794 CTCTGCAATGGCCTTGAGGGCTTTGCTGCAGCCACCCCTCCTCCCAGTTCCACTTGCAAG 853
              |||||||||||||||||||||||||||||||||||  |  ||||||||||||||||||
Db        658 CTCTGCAATGGCCTTGAGGGCTTTGCTGCAGCCACCCCCTCCTCCAGTTCCACTTGCAAG 717

Qy        854 TCTGACCTGGCTGAGCTGGACCATGTGGTAGAGATTCTGGT-GGAGACCTGAGAGGCCAC 912
              ||||||||| |||||||||||||||||||||||||||||| ||||||||||||||||||
Db        718 TCTGACCTGNCTGAGCTGGACCATGTGGTAGAGATTCTGGTGGGAGACCTGAGAGGCCAC 777

Qy        913 CCCAGT-GGGCTAAGGGTGAGGCCACCAGTCCCCATGGAGCTCACGTGTGTTGTGACCCA 971
              |||||| ||||||| ||||||||||||||||||||||||||||||||||||  |||
Db        778 CCCAGTGGGGCTAAAGGTGAGGCCACCAGTCCCCATGGAGCTCACGTGTGTTGTGGACCA 837

Qy        972 GAGACAGATAAGCACTTGTCCTAAGA--GGGGCTCTGGCTCTTGAGCTCATT-ATCCTTT 1028
              |||||||||||||||||||||  |||  |||      |||||||||||||||| |||||||
Db        838 GAGACAGATAAGCACTTGTCCCTAGAAGGGGCTCTGGGCTCTTGAGCTCATTAATCCTTT 897

Qy       1029 TGTGT-GACATTGGACTCACTG 1049
              ||||| |||||||||||| |||
Db        898 TGTGTNGACATTGGACTCCCTG 919


RESULT 10
BB609465
LOCUS       BB609465                994 bp    mRNA    linear   EST 31-AUG-2001
DEFINITION  BB609465 RIKEN full-length enriched, adult male kidney Mus musculus
            cDNA clone 0610010F24 5', mRNA sequence.
ACCESSION   BB609465
VERSION     BB609465.1  GI:15390279
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 994)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T.,
            Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J.,
            Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K.,
            Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
```

Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Yamanaka,I., Kiyosawa,H., Kondo,S., Saito,T., Shinagawa,A., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K., Arakawa,T., Ishii,Y. and Hayashizaki,Y.
Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct. Func. Genomics 2 pre, L72-L86 (2001
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.

FEATURES Location/Qualifiers
source 1. .994
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="0610010F24"
/sex="male"
/tissue_type="kidney"
/dev_stage="adult"
/lab_host="SOLR"
/clone_lib="RIKEN full-length enriched, adult male kidney"
/note="Site_1: XhoI; Site_2: SstI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGAAGGATCCAAGAGCTCAATTAATTAATTAAACCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. "

ORIGIN

Query Match 36.3%; Score 755.4; DB 10; Length 994;
Best Local Similarity 96.2%; Pred. No. 2.7e-140;
Matches 815; Conservative 0; Mismatches 23; Indels 9; Gaps 4;

```
Qy        182 CTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGAAGGAGTA 241
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 CCGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGAAGGAGTA 142

Qy        242 GAGGGTTTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACAGTCACTCCTGGACATGTCC 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 GAGGGTTTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACAGTCACTCCTGGACATGTCC 202

Qy        302 CTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAATCTCTGCCGCTCGGTCCTC 361
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        203 CTTGTCAAGCTCCAGCTCTGTCACGTGCTAGTGGAGCCCAATCTCTGCCGCTCGGTCCTC 262

Qy        362 ATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCAGGATGGTGTGTGGCATGGG 421
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 ATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCAGGATGGTGTGTGGCATGGG 322

Qy        422 ATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCTGGTGTCCACAGAGATCCTG 481
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        323 ATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCTGGTGTCCACAGAGATCCTG 382

Qy        482 TGTCGTACAGTGAGGGGAGCTGAGGAAGAGCACCCTGCTCCTGAACTGGAAGATGCTCCC 541
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 TGTCGTACAGTGAGGGGAGCTGAGGAAGAGCACCCTGCTCCTGAACTGGAAGATGCTCCC 442

Qy        542 TTGCAAAACTCGGTTTCCGAGCTCCCCATCGTTGGCTCAGCACCAGGGCAAAGGAACCCT 601
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        443 TTGCAAAACTCGGTTTCCGAGCTCCCCATCGTTGGCTCAGCACCAGGGCAAAGGAACCCT 502

Qy        602 CAGAGCAGCCTCTGGGAGATGGACAGCCCACAAGAAAACAGGGGAAGCTTTCAGAAGTCA 661
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        503 CAGAGCAGCCTCTGGGAGATGGACAGCCCACAAGAAAACAGGGGAAGCTTTCAGAAGTCA 562

Qy        662 CTGGACCAGATATTTGAGACCCTGGAGAACAAAAACTCCAGTTCAGTGGAGGAACTCTTC 721
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        563 CTGGACCAGATATTTGAGACCCTGGAGAACAAAAACTCCAGTTCAGTGGAGGAACTCTTC 622

Qy        722 TCAGATGTGGACAGCTCCTACTATGACCTGGACACAGTGCTAACAGGAATGATGAGTGGG 781
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        623 TCAGATGTGGACAGCTCCTACTATGACCTGGACACAGTGCTAACAGGAATGATGAGTGGG 682

Qy        782 ACCAAGTCCAGTCTCTGCAATGGCCTTGAGGGCTTTGCTGCAGCCACCCCTCCTCCCAGT 841
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        683 ACCAAGTCCAGTCTCTGCAATGGCCTTGAGGGCTTTGCTGCAGCCACCCCTCCTCCCAGT 742

Qy        842 TCCACTTGCAAGTCTGACCTGGCTGAGCTGGACCATGTGGTAGAGATTCTGGTGGAGACC 901
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        743 TCCACTTGCAAGTCTGACCTGGCTGAGCTGGACCATGTGGTAGAGATTCTGGTGGAGACC 802

Qy        902 TGAGAGGCCACCCCAGTGGGCTAAGGGTGAGGCCACCAGTCCCCATGGAGCTCACGTGTG 961
              ||||||| ||||||||||||||||||||||||| ||    |||||||||| |||||||
Db        803 TGAGAGGTCACCCCAGTGGGCTAAGGGTGAGGC--CCCAGTCCCATGGAGCT-ACGTGTG 859

Qy        962 TTGTGACCCAGAGACAGATAAGCACTTGTCCTAAGAGGGGCTCTGGCTCTTGAGCTCATT 1021
              | ||||||   ||| |    | ||||||||||   ||| ||  |   ||||||||||
Db        860 TAGTGACC---AGAGAAGATAACACTTGTCCTAGAGGGGTCT---GGCTCTGAGCTCATT 913

Qy       1022 ATCCTTT 1028
              ||| |||
Db        914 ATCTTTT 920
```

RESULT 11
BQ936878
LOCUS BQ936878 967 bp mRNA linear EST 21-AUG-2002
DEFINITION AGENCOURT_8930019 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:6488699 5', mRNA sequence.
ACCESSION BQ936878
VERSION BQ936878.1 GI:22352261
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 967)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM14035 row: 1 column: 12
High quality sequence stop: 635.

```
FEATURES             Location/Qualifiers
     source          1. .967
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N-3"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6488699"
                     /tissue_type="tumor, biopsy sample"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Mam2"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
ORIGIN

  Query Match           35.9%;  Score 748.4;  DB 13;  Length 967;
  Best Local Similarity   96.4%;  Pred. No. 6.9e-139;
  Matches  787;  Conservative    0;  Mismatches   26;  Indels    3;  Gaps    2;

Qy        215 AAATATGGTGACCAGGAAGAAGGAGTAGAGGGTTTTGGCACTGTCCCTTCCTATAGCCTG 274
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAATATGGTGACCAGGAAGAAGGAGTAGAGGGTTTTGGCACTGTCCCTTCCTATAGCCTG 60

Qy        275 CAGCGACAGTCACTCCTGGACATGTCCCTTGTCAAGCTCCAGCTCTGTCACATGCTAGTG 334
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAGCGACAGTCACTCCTGGACATGTCCCTTGTCAAGCTCCAGCTCTGTCACATGCTAGTG 120

Qy        335 GAGCCCAATCTCTGCCGCTCGGTCCTCATCGCCAACACAGTCCGGCAGATCCAGGAGGAA 394
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GAGCCCAATCTCTGCCGCTCGGTCCTCATCGCCAACACAGTCCGGCAGATCCAGGAGGAA 180

Qy        395 ATGAGCCAGGATGGTGTGTGGCATGGGATGGCACCCCAGAATGTAGATCGGGCACCAGTT 454
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATGAGCCAGGATGGTGTGTGGCATGGGATGGCACCCCAGAATGTAGATCGGGCACCAGTT 240

Qy        455 GAACGCCTGGTGTCCACAGAGATCCTGTGTCGTACAGTGAGGGGAGCTGAGGAAGAGCAC 514
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAGCGCCTGGTGTCCACAGAGATCCTGTGTCGTACAGTGAGGGGAGCTGAGGAAGAGCAC 300

Qy        515 CCTGCTCCTGAACTGGAAGATGCTCCCTTGCAAAACTCGGTTTCCGAGCTCCCCATCGTT 574
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CCTGCTCCTGAACTGGAAGATGCTCCCTTGCAAAACTCGGTTTCCGAGCTCCCCATCGTT 360

Qy        575 GGCTCAGCACCAGGGCAAAGGAACCCTCAGAGCAGCCTCTGGGAGATGGACAGCCCACAA 634
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCTCAGCACCAGGGCAAAGGAACCCTCAGAGCAGCCTCTGGGAGATGGACAGCCCACAA 420

Qy        635 GAAAACAGGGGAAGCTTTCAGAAGTCACTGGACCAGATATTTGAGACCCTGGAGAACAAA 694
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GAAAACAGGGGAAGCTTTCAGAAGTCACTGGACCAGATATTTGAGACCCTGGAGAACAAA 480

Qy        695 AACTCCAGTTCAGTGGAGGAACTCTTCTCAGATGTGGACAGCTCCTACTATGACCTGGAC 754
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AACTCCAGTTCAGTGGAGGAACTCTTCTCAGATGTGGACAGCTCCTACTATGACCTGGAC 540

Qy        755 ACAGTGCTAACAGGAATGATGAGTGGGACCAAGTCCAGTCTCTGCAATGGCCTTGAGGGC 814
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        541 ACAGTGCTAACAGGAATGATGACTGGGACCAAGTCCAGTCTCTGCAATGGCCTTGAGGGC 600

Qy        815 TTTGCTGCAGCCACCCCTCCTCCCAGTTCCACTTGCAAGTCTGACCTGGCTGAGCTGGAC 874
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTTGCTGCAGCCGCCCCTCCTCCCAGTTCCACTTGCAAGTCTGACCTGGCTGAGCTGGAC 660

Qy        875 CATGTGGTAGAGATTCTGGTGGAGACCTGAGAGGCCACCCCAGTGGGCTAAGGGTGAGGC 934
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        661 CATGTGGTAGAGATTCTGGTGGAGACCTGAGAGGCCCCCCCAGTGGGCTAAGGGTGAGGT 720

Qy        935 CACCAGTCCCCATGG-AGCTCACGTGTGTTGTGACCCAGAGACAGATAAGCACTT--GTC 991
              |||||| |||||||| | |||||||||||||||||    | || ||||| |||||     |
Db        721 CACCAGGCCCCATGGAACCTCACGTGTGTTGTGACCGCCACACCGATAAACACTTTGTCC 780

Qy        992 CTAAGAGGGGCTCTGGCTCTTGAGCTCATTATCCTT 1027
                |  | |||||| ||||||||| |  ||||| | |
Db        781 TAAGAACGGGCTCCGGCTCTTGACCATATTATACCT 816


RESULT 12
BQ713776
LOCUS       BQ713776                922 bp    mRNA    linear   EST 16-JUL-2002
DEFINITION  AGENCOURT_8295059 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6309426
            5', mRNA sequence.
ACCESSION   BQ713776
VERSION     BQ713776.1  GI:21852675
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 922)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Susan L. Sullivan, PhD.
             cDNA Library Preparation: ResGen, Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13730  row: b  column: 19
            High quality sequence stop: 709.
FEATURES             Location/Qualifiers
     source          1. .922
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6309426"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_129"
                     /note="Organ: olfactory epithelium; Vector: pCMV-SPORT6.1;
                     Site_1: EcoRV; Site_2: NotI; Cloned unidirectionally.
                     Primer: Oligo dT. Average insert size 2.2 kb. Constructed
                     by ResGen, Invitrogen Corp. Note: this is a NIH_MGC
                     Library."
ORIGIN

  Query Match           35.8%;  Score 746;  DB 13;  Length 922;
  Best Local Similarity   95.9%;  Pred. No. 2.1e-138;
  Matches  822;  Conservative    0;  Mismatches   21;  Indels   14;  Gaps    5;

Qy        134 TCCGGGTGCCCGAAGGAGCTGGCGTGGGTCTGCCTTGCAGCCGCCCGCCTGGACAGGATG 193
              |||||||||||||||||   |   |||||||||||| ||       | ||||||||||
Db         25 TCCGGGTGCCCGAAGGAG--TGGCGGGGTCTGCCTTGCGGC------GCCGGACAGGATG 76

Qy        194 TTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGAAGGAGTAGAGGGTTTTGGC 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         77 TTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGAAGGAGTAGAGGGTTTTGGC 136

Qy        254 ACTGTCCCTTCCTATAGCCTGCAGCGACAGTCACTCCTGGACATGTCCCTTGTCAAGCTC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        137 ACTGTCCCTTCCTATAGCCTGCAGCGACAGTCACTCCTGGACATGTCCCTTGTCAAGCTC 196

Qy        314 CAGCTCTGTCACATGCTAGTGGAGCCCAATCTCTGCCGCTCGGTCCTCATCGCCAACACA 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        197 CAGCTCTGTCACATGCTAGTGGAGCCCAATCTCTGCCGCTCGGTCCTCATCGCCAACACA 256

Qy        374 GTCCGGCAGATCCAGGAGGAAATGAGCCAGGATGGTGTGTGGCATGGGATGGCACCCCAG 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        257 GTCCGGCAGATCCAGGAGGAAATGAGCCAGGATGGTGTGTGGCATGGGATGGCACCCCAG 316

Qy        434 AATGTAGATCGGGCACCAGTTGAACGCCTGGTGTCCACAGAGATCCTGTGTCGTACAGTG 493
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        317 AATGTAGATCGGGCACCAGTTGAACGCCTGGTGTCCACAGAGATCCTGTGTCGTACAGTG 376

Qy        494 AGGGGAGCTGAGGAAGAGCACCCTGCTCCTGAACTGGAAGATGCTCCCTTGCAAAACTCG 553
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        377 AGGGGAGCTGAGGAAGAGCACCCTGCTCCTGGACTGGAAGATGCTCCCTTGCAAAACTCG 436

Qy        554 GTTTCCGAGCTCCCCATCGTTGGCTCAGCACCAGGGCAAAGGAACCCTCAGAGCAGCCTC 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        437 GTTTCCGAGCTCCCCATCGTTGGCTCAGCACCAGGGCAAAGGAACCCTCAGAGCAGCCTC 496

Qy        614 TGGGAGATGGACAGCCCACAAGAAAACAGGGGAAGCTTTCAGAAGTCACTGGACCAGATA 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        497 TGGGAGATGGACAGCCCACAAGAAAACAGGGGAAGCTTTCAGAAGTCACTGGACCAGATA 556

Qy        674 TTTGAGACCCTGGAGAACAAAAACTCCAGTTCAGTGGAGGAACTCTTCTCAGATGTGGAC 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        557 TTTGAGACCCTGGAGAACAAAAACTCCAGTTCAGTGGAGGAACTCTTCTCAGATGTGGAC 616

Qy        734 AGCTCCTACTATGACCTGGACACAGTGCTAACAGGAATGATGAGTGGGACCAAGTCCAGT 793
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        617 AGCTCCGACTATGACCTGGACACAGTGCTAACAGGAATGATGAGTGGGACCAAGTCCAGT 676

Qy        794 CTCTGCAATGGCCTTGAGGGCTTTGCTGCAGCCACCCCTCCTCCCAGTTCCACTTGCAAG 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        677 CTCTGCAATGGCCTTGAGGGCTTTGCTGCAGCCACCCCTCCTCCCAGTTCCACTTGCAAG 736

Qy        854 TCTGACCTGGCTGAGCTGGACCATGTGGTAGAGATTCTGGT-GGAGACCTGAGAGGCCAC 912
              |||||||||||||||||||| |||||||||||||||||| |||||||||||||||||||
Db        737 TCTGACCTGGCTGAGCTGGACCGTGTGGTAGAGATTCTGGTGGGAGACCTGAGAGGCCAC 796

Qy        913 CCCAGTGGGCTAAGGGTGAGGCCA-CCAGTCCCCATGGAGCTCACGTGTGTTGT----GA 967
              |||||||||||||||||||||||| |||||||||||||      |  |||| ||
Db        797 CCCAGTGGGCTAAGGGTGAGGCCACCCAGTCCCCATGGNAGCTCCACGTGTGGTTGTGAC 856

Qy        968 CCCAGAGACAGATAAGC 984
              |||||||||||||||||
Db        857 CCCAGAGACAGATAAGC 873


RESULT 13
BI661691
LOCUS       BI661691                 830 bp    mRNA    linear   EST 12-SEP-2001
DEFINITION  603305632F1 NCI_CGAP_Mam4 Mus musculus cDNA clone IMAGE:5351646 5',
            mRNA sequence.
ACCESSION   BI661691
VERSION     BI661691.1  GI:15575927
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 830)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth
            Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11893  row: o  column: 07
            High quality sequence stop: 797.
FEATURES             Location/Qualifiers
     source          1. .830
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="NMRI"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5351646"
                     /tissue_type="tumor, gross tissue"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Mam4"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigators
                     providing samples: Lothar Hennighausen/Priscilla Furth,
                     NIH Reference for transgenic model: Li et al., Cell Growth
                     and  Differentiation 7, 3-11 (1996)."
ORIGIN

  Query Match           35.6%;  Score 741;  DB 12;  Length 830;
  Best Local Similarity   98.7%;  Pred. No. 2.1e-137;
  Matches  755;  Conservative    0;  Mismatches    9;  Indels    1;  Gaps    1;

Qy        182 CTGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGAAGGAGTA 241
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 CCGGACAGGATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGAAGGAGTA 125

Qy        242 GAGGGTTTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACAGTCACTCCTGGACATGTCC 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        126 GAGGGTTTTGGCACTGTCCCTTCCTATAGCCTGCAGCGACAGTCACTCCTGGACATGTCC 185

Qy        302 CTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAATCTCTGCCGCTCGGTCCTC 361
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        186 CTTGTCAAGCTCCAGCTCTGTCACATGCTAGTGGAGCCCAATCTCTGCCGCTCGGTCCTC 245

Qy        362 ATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCAGGATGGTGTGTGGCATGGG 421
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        246 ATCGCCAACACAGTCCGGCAGATCCAGGAGGAAATGAGCCAGGATGGTGTGTGGCATGGG 305

Qy        422 ATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAACGCCTGGTGTCCACAGAGATCCTG 481
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        306 ATGGCACCCCAGAATGTAGATCGGGCACCAGTTGAGCGCCTGGTGTCCACAGAGATCCTG 365

Qy        482 TGTCGTACAGTGAGGGGAGCTGAGGAAGAGCACCCTGCTCCTGAACTGGAAGATGCTCCC 541
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        366 TGTCGTACAGTGAGGGGAGCTGAGGAAGAGCACCCTGCTCCTGAACTGGAAGATGCTCCC 425

Qy        542 TTGCAAAACTCGGTTTCCGAGCTCCCCATCGTTGGCTCAGCACCAGGGCAAAGGAACCCT 601
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        426 TTGCAAAACTCGGTTTCCGAGCTCCCCATCGTTGGCTCAGCACCAGGGCAAAGGAACCCT 485

Qy        602 CAGAGCAGCCTCTGGGAGATGGACAGCCCACAAGAAAACAGGGGAAGCTTTCAGAAGTCA 661
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        486 CAGAGCAGCCTCTGGGAGATGGACAGCCCACAAGAAAACAGGGGAAGCTTTCAGAAGTCA 545

Qy        662 CTGGACCAGATATTTGAGACCCTGGAGAACAAAAACTCCAGTTCAGTGGAGGAACTCTTC 721
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        546 CTGGACCAGATATTTGAGACCCTGGAGAACAAAAACTCCAGTTCAGTGGAGGAACTCTTC 605

Qy        722 TCAGATGTGGACAGCTCCTACTATGACCTGGACACAGTGCTAACAGGAATGATGAGTGGG 781
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db        606 TCAGATGTGGACAGCTCCTACTATGACCTGGACACAGTGCTAACAGGAATGATGACTGGG 665

Qy        782 ACCAAGTCCAGTCTCTGCAATGGCCTTGAGGGCTTTGCTGCAGCCACCCCTCCTCCCAGT 841
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db        666 ACCAAGTCCAGTCTCTGCAATGGCCTTGAGGGCTTTGCTGCAGCCGCCCCTCCTCCCAGT 725

Qy        842 TCCACTTGCAAGTCTGACCTGGCTGAGCTGGACCATGTGGTAGAGATTCTGGTGGAGACC 901
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db        726 TCCACTTGCAAGTCTGACCTGGCTGAGCTGGAACATGTGGTAGAGATTCTGGTGGAGACC 785

Qy        902 TGAGAGGCCACCCCAGTGGGCTAAGGGTGAGGCCACCAGTCCCCA 946
              ||||||| ||||| ||||||| |||||||| | ||||||||||||
Db        786 TGAGAGG-CACCCAAGTGGGCAAAGGGTGAAGGCACCAGTCCCCA 829


RESULT 14
BI153019
LOCUS       BI153019                 800 bp    mRNA    linear   EST 05-JUL-2001
```

```
DEFINITION  602918211F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:5068418 5',
            mRNA sequence.
ACCESSION   BI153019
VERSION     BI153019.1  GI:14613020
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 800)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11184  row: f  column: 03
            High quality sequence stop: 797.
FEATURES             Location/Qualifiers
     source          1. .800
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="Czech II"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5068418"
                     /tissue_type="spontaneous tumor, metastatic to mammary.
                     Stem cell origin."
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Lu29"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
ORIGIN
```

Query Match 34.6%; Score 721.4; DB 12; Length 800;
Best Local Similarity 98.3%; Pred. No. 1.8e-133;
Matches 770; Conservative 0; Mismatches 8; Indels 5; Gaps 4;

```
Qy       1270 TGGTGTGTCATTTTAAAGGTTTAAATAGAAAACTTCTTTTGTTACTGAGTCTCTACACTC 1329
             ||||||||||||| || |||||||||||||||||| |||||||||||||||||||||||||
Db         21 TGGTGTGTCATTTGAAGGGTTTAAATAGAAAACTTC-TTTGTTACTGAGTCTCTACACTC 79

Qy       1330 CCAAGGCAACTGTAAATGTAGCCGGCCGGGTGTTTACATGAGAGGCTCCAGTATGGTCTA 1389
             |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db         80 CCAAGGCAACTGTAAATGTAGCTGGCCGGGTGTTTACATGAGAGGCTCCAGTATGGTCTA 139

Qy       1390 CATTCTAGTAGAGCTTGAAA--AGAACCATGCACAGCTCCACTGCCCCCTCACTGGGTCT 1447
             |||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db        140 CATTCTAGTAGAGCTTGAAAAGAGAACCATGCACAGCTCCACTGCCCCCTCACTGGGTCT 199

Qy       1448 GCTCTGGCGGATCGGAGCTCTCTTCCTAGCCCCGTGTGCAGGATGGCTTTATTTATGCCT 1507
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        200 GCTCTGGCGGATCGGAGCTCTCTTCCTAGCCCCGTGTGCAGGATGGCTTTATTTATGCCT 259

Qy       1508 ATTTATATGTAAATGCCACTGAAAGCTAAGGTCTTACTCCTGGAAATCCCAACACCAGTT 1567
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        260 ATTTATATGTAAATGCCACTGAAAGCTAAGGTCTTACTCCTGGAAATCCCAACACCAGTT 319

Qy       1568 CTTCAGGGACTGCTGTGAGGCAGTGCCTTATGCAGGTCTTGTCCTTGGCCATCACTGTCT 1627
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        320 CTTCAGGGACTGCTGTGAGGCAGTGCCTTATGCAGGTCTTGTCCTTGGCCATCACTGTCT 379

Qy       1628 GGTTCCCAGCCCAGCACATGTGACATGAGGACATGACATGCCCGAACCACCCAGCACCAC 1687
             |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        380 GGTTCCCAGCCCAGCACATGTGACATGAGGACATGACATGCCCGAACCA-CCAGCACCAC 438

Qy       1688 ATGCTCCATGTCAAGTGTGTACGTGGAGACCACTGGCTCCCAGGCCTGTGCTCAGAGAGG 1747
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        439 ATGCTCCATGTCAAGTGTGTACGTGGAGACCACTGGCTCCCAGGCCTGTGCTCAGAGAGG 498

Qy       1748 GTGTGCAGTCCTACGTGTGCTGGGGGGGACGACGGTGACCTGTGCTTGCTTGCTTTTAAA 1807
             |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        499 GTGTGCAGTCCCACGTGTGCTGGGGGGGACGACGGTGACCTGTGCTTGCTTGCTTTTAAA 558

Qy       1808 ATGGTGCTTGGACGTTTTAAGGTTAAAAACAATCCGACTCCATATGATTTAGGGCTCCTC 1867
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        559 ATGGTGCTTGGACGTTTTAAGGTTAAAAACAATCCGACTCCATATGATTTAGGGCTCCTC 618

Qy       1868 CACCCTGGGGTGGCCCCTATGCTGTCTGCTTGGATCTCAAAGTCTTGGTACTCGGCACTG 1927
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        619 CACCCTGGGGTGGCCCCTATGCTGTCTGCTTGGATCTCAAAGTCTTGGTACTCGGCACTG 678

Qy       1928 TCAGACTCCACCCCATGTATCCTTTTTGTTTCTCTTGTGCTTTTTTTGGACTTCCCAACC 1987
             |||||||||| ||| ||||||||||||||||||||||||||||||||||||||||  |
Db        679 TCAGACTCCANCCCGTGTATCCTTTTTGTTTCTCTTGTGCTTTTTTTGGACTTCCCANAC 738

Qy       1988 TGAGCCTAAGGTTTTATTTTATATGTGCTTCAATATCAACAATGTAAACCTCACTTTATT 2047
             |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        739 TGAGCCTAAGG-TTTATTTTATATGTGCTTCAATATCAACAATGTAAACCTCACTTTATT 797

Qy       2048 AAA 2050
             |||
Db        798 AAA 800
```

```
RESULT 15
AY403475
LOCUS       AY403475                714 bp    DNA     linear   GSS 15-DEC-2003
DEFINITION  Mus musculus CDCA4 gene, VIRTUAL TRANSCRIPT, partial sequence,
            genomic survey sequence.
ACCESSION   AY403475
VERSION     AY403475.1  GI:39759458
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 714)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Inferring nonneutral evolution from human-chimp-mouse orthologous
            gene trios
  JOURNAL   Science 302 (5652), 1960-1963 (2003)
   PUBMED   14671302
REFERENCE   2  (bases 1 to 714)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT     This sequence was made by sequencing genomic exons and ordering
            them based on alignment.
FEATURES             Location/Qualifiers
     source          1. .714
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
     gene            <1. .>714
                     /gene="CDCA4"
                     /locus_tag="HCM1569"
ORIGIN
```

Query Match 34.3%; Score 714; DB 29; Length 714;
Best Local Similarity 100.0%; Pred. No. 5.4e-132;
Matches 714; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        191 ATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGAAGGAGTAGAGGGTTTT 250
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTTTGCTAGAGGGCTGAAGAGGAAATATGGTGACCAGGAAGAAGGAGTAGAGGGTTTT 60

Qy        251 GGCACTGTCCCTTCCTATAGCCTGCAGCGACAGTCACTCCTGGACATGTCCCTTGTCAAG 310
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGCACTGTCCCTTCCTATAGCCTGCAGCGACAGTCACTCCTGGACATGTCCCTTGTCAAG 120

Qy        311 CTCCAGCTCTGTCACATGCTAGTGGAGCCCAATCTCTGCCGCTCGGTCCTCATCGCCAAC 370
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTCCAGCTCTGTCACATGCTAGTGGAGCCCAATCTCTGCCGCTCGGTCCTCATCGCCAAC 180

Qy        371 ACAGTCCGGCAGATCCAGGAGGAAATGAGCCAGGATGGTGTGTGGCATGGGATGGCACCC 430
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACAGTCCGGCAGATCCAGGAGGAAATGAGCCAGGATGGTGTGTGGCATGGGATGGCACCC 240

Qy        431 CAGAATGTAGATCGGGCACCAGTTGAACGCCTGGTGTCCACAGAGATCCTGTGTCGTACA 490
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGAATGTAGATCGGGCACCAGTTGAACGCCTGGTGTCCACAGAGATCCTGTGTCGTACA 300

Qy        491 GTGAGGGGAGCTGAGGAAGAGCACCCTGCTCCTGAACTGGAAGATGCTCCCTTGCAAAAC 550
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GTGAGGGGAGCTGAGGAAGAGCACCCTGCTCCTGAACTGGAAGATGCTCCCTTGCAAAAC 360

Qy        551 TCGGTTTCCGAGCTCCCCATCGTTGGCTCAGCACCAGGGCAAAGGAACCCTCAGAGCAGC 610
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCGGTTTCCGAGCTCCCCATCGTTGGCTCAGCACCAGGGCAAAGGAACCCTCAGAGCAGC 420

Qy        611 CTCTGGGAGATGGACAGCCCACAAGAAAACAGGGGAAGCTTTCAGAAGTCACTGGACCAG 670
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTCTGGGAGATGGACAGCCCACAAGAAAACAGGGGAAGCTTTCAGAAGTCACTGGACCAG 480

Qy        671 ATATTTGAGACCCTGGAGAACAAAAACTCCAGTTCAGTGGAGGAACTCTTCTCAGATGTG 730
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATATTTGAGACCCTGGAGAACAAAAACTCCAGTTCAGTGGAGGAACTCTTCTCAGATGTG 540

Qy        731 GACAGCTCCTACTATGACCTGGACACAGTGCTAACAGGAATGATGAGTGGGACCAAGTCC 790
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GACAGCTCCTACTATGACCTGGACACAGTGCTAACAGGAATGATGAGTGGGACCAAGTCC 600

Qy        791 AGTCTCTGCAATGGCCTTGAGGGCTTTGCTGCAGCCACCCCTCCTCCCAGTTCCACTTGC 850
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AGTCTCTGCAATGGCCTTGAGGGCTTTGCTGCAGCCACCCCTCCTCCCAGTTCCACTTGC 660

Qy        851 AAGTCTGACCTGGCTGAGCTGGACCATGTGGTAGAGATTCTGGTGGAGACCTGA 904
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AAGTCTGACCTGGCTGAGCTGGACCATGTGGTAGAGATTCTGGTGGAGACCTGA 714
```

Search completed: September 10, 2004, 14:17:08
Job time : 5672 secs